بسم‌الله‌الرحمن‌الرحیم

والحمدلله القاصم الجبارین

اندیشکده قرار

تبدیل تهدیدها به فرصت‌ها یکی از اهداف اصلی مقاومت است. فهمیدن این فرصتی که در دل تهدیدها نهفته است، نیاز به مطالعاتی علمی، روش‌مند و در عین‌حال هدفمند و با دغدغه دارد. مطالعاتی ساختاریافته که به دنبال رسیدن به هدفی مشخص در میدان باشد، از پرداختن بیش از اندازه به کلیات و تحلیل‌های بدون مستندات دقیق اجتناب کند، محدودیت‌های میدانی را بشناسد و از اطناب آکادمیک اجتناب کند. اندیشکده قرار تلاش دارد این نیاز حیاتی جبهه مقاومت را برطرف سازد و به محلی برای همگرایی مطالعات ساختاریافته و هدفمند در زمینه مقاومت و استعمار در منطقه غرب آسیا تبدیل شود.

هوش مصنوعی در خدمت جنگ

بررسی ابعاد فنی و حقوقی استفاده از سامانه «لاوندر» در جنگ غزه

از اکتبر ۲۰۲۳، جنگ غزه شاهد یکی از نخستین نمونه‌های استفاده گسترده از هوش مصنوعی در عملیات نظامی بوده است. ارتش اسرائیل با به‌کارگیری سامانه‌ای موسوم به لاوندر (Lavender) کوشید فرآیند شناسایی و هدف‌گیری افراد را خودکار کند. این سامانه با بهره‌گیری از یادگیری ماشینی و تشخیص چهره، داده‌های مکانی، ارتباطی و تصویری را پردازش می‌کند تا فهرستی از اهداف احتمالی تولید کند.

گزارش‌ها نشان می‌دهد که در هفته‌های نخست جنگ، لاوندر ده‌ها هزار "هدف بالقوه" را شناسایی کرده و بسیاری از حملات هوایی بر پایه این فهرست انجام شده است. این تحول، نقطه‌ی عطفی در مفهوم « مداخله ماشین در فرآیند تصمیم‌گیری در طی جنگ » محسوب می‌شود و پرسش‌هایی اساسی درباره قابلیت اطمینان، سوگیری، مسئولیت انسانی و انطباق این فرآیند با اصول حقوق بین‌الملل بشردوستانه (IHL) مطرح کرده است.

در همین راستا، امیلی اندرسن (Emelie Andersin)، پژوهشگر دکتری رشته مدیریت و روابط بین‌الملل[1] دانشگاه لیدن هلند، در مقاله‌ای با عنوان استفاده از سامانه "لاوندر" در غزه و قانون انتخاب اهداف در عملیات نظامی: سامانه‌های پشتیبان تصمیم‌گیری هوش‌مصنوعی و فناوری تشخیص چهره « The Use of the ‘Lavender’ in Gaza and the Law of Targeting: ai-Decision Support Systems and Facial Recognition Technology»، به بررسی کاربست ابزارهای هوش مصنوعی از منظر حقوق بین‌الملل بشردوستانه (IHL) و اصول اخلاقی جنگ پرداخته است. او در این مقاله با تحلیل نمونه سامانه لاوندر، نشان می‌دهد که چگونه انتقال تصمیم‌گیری از انسان به ماشین، اصول بنیادین حقوق بین‌الملل بشردوستانه را به چالش می‌کشد و موجب شکل‌گیری «خلاء مسئولیت‌پذیری (Accountability Gap)» در فرآیندهای نظامی می‌شود.

به باور اندرسن، مسئله اصلی در استفاده از سامانه‌هایی چون لاوندر، نه صرفاً در خطاهای فنی، بلکه در غیبت کنترل انسانی مؤثر و شفافیت نحوه تصمیم‌گیری است. این سامانه‌ها برمبنای مدل‌های احتمالاتی طراحی شده‌اند که همواره سطحی از عدم قطعیت را با خود به همراه دارند که در محیطی مانند غزه، همین عدم قطعیت می‌تولند به قیمت جان هزاران غیرنظامی تمام شود. افزون‌بر این، سوگیری داده‌ها، معماری غیرشفاف و اتکای بیش از حد به خروجی ماشین، سه چالش اساسی هستند که از منظر اندرسن، سامانه‌های هوش مصنوعی را از تطبیق با اصول سه‌گانه قانون انتخاب اهداف در عملیات نظامی - تمایز، تناسب و احتیاط- بازمی‌دارند.

از دید اندرسن، سامانه لاوندر به‌طور ساختاری با اصل تمایز در تضاد است. این اصل دولت‌ها را ملزم می‌کند میان اهداف نظامی و غیرنظامی تمایز قائل شوند، اما لاوندر با اتکا بر الگوهای آماری رفتاری، معیارهای انسانی تمایز را به احتمالات ریاضی تقلیل داده است. در چنین چارچوبی، فردی ممکن است تنها به دلیل شباهت آماری رفتارهای روزمره‌اش با اعضای یک گروه مسلح، به عنوان هدف مشروع طبقه‌بندی شود. اندرسن تأکید می‌کند که وقتی الگوریتم جایگزین قضاوت انسانی در تشخیص هدف می‌شود، مفهوم «مشکوک بودن» دیگر مبتنی‌بر شواهد عینی نیست، بلکه صرفاً خروجی محاسبه‌ای غیرقابل‌توضیح است. از این رو، عملکرد لاوندر عملاً مرز میان هدف نظامی و غیرنظامی را از میان برده و اصل تمایز را بی‌اثر کرده است.

همچنین از نگاه او، اصل تناسب که می‌کوشد میان مزیت نظامی و خسارت غیرنظامی توازن برقرار کند، در فضای الگوریتمی تقریباً بی‌معنا می‌شود؛ زیرا مزیت نظامی را نمی‌توان مانند تلفات غیرنظامیان به‌صورت کمی سنجید. در نتیجه، هرگونه

[1] Governance and Global Affairs

«محاسبه» میان این دو، فاقد پشتوانه منطقی و انسانی است. به بیان دیگر، وقتی تصمیم هدف‌گیری از انسان به ماشین سپرده شود، «موازنه اخلاقی» از میان می‌رود.

اندِرسن اصل احتیاط را نیز از بنیادی‌ترین اصولی می‌داند که با اتوماسیون نظامی در تعارض قرار گرفته است. مطابق ماده ۵۷ پروتکل الحاقی اول کنوانسیون ژنو، فرماندهان موظف‌اند پیش از هر حمله، تمامی تدابیر ممکن را برای اطمینان از مشروعیت هدف و کاهش خسارات غیرنظامی اتخاذ کنند. اما در سامانه لاوندر، زمان ارزیابی هر هدف به‌طور متوسط کمتر از ۲۰ ثانیه است، و نقش انسان از تصمیم‌گیرنده به صرف تأییدکننده تقلیل یافته است. اندرسن این پدیده را نمونه‌ای از «سوگیری اتوماسیون» می‌داند که به تمایل کاربران به پذیرش خروجی ماشین بدون بررسی انتقادی اشاره دارد و عملاً اصل احتیاط را از بین می‌برد. او هشدار می‌دهد که در چنین سیستمی، تصمیم درباره مرگ و زندگی انسان‌ها به تابعی از محاسبات آماری بدل شده و نظارت اخلاقی به حداقل رسیده است.

اندِرسن در جمع‌بندی خود تأکید می‌کند که ورود هوش مصنوعی به میدان جنگ، بازتعریف رابطه‌ی انسان، اخلاق و فناوری را ضروری می‌سازد. بدون وجود شفافیت، قابلیت توضیح‌پذیری و کنترل انسانی مؤثر، چنین سامانه‌هایی نه ابزاری برای افزایش دقت، بلکه عامل گسترش خطا و کشتار در مقیاس صنعتی خواهند بود. تحلیل اندرسن نه‌تنها هشدار نسبت به آینده‌ی جنگ‌های الگوریتمی است، بلکه تلاشی است برای بازگرداندن انسان به مرکز تصمیم‌گیری در جهانی که روزبه‌روز بیشتر به ماشین‌ها می‌سپارد.

اما موضوع مهمی که در این گزارش به آن پرداخته نشده است این است که سامانه‌های هوش مصنوعی در جنگ غزه، نه صرفاً ابزار نظامی، بلکه بخشی از سازوکار مشروعیت‌بخشی به سیاست پاکسازی نژادی اسرائیل بوده‌اند. استفاده از سامانه‌هایی نظیر لاوندر، که تصمیم‌گیری درباره حیات انسان‌ها را به مدل‌های آماری می‌سپارد، فراتر از یک تحول فناورانه است؛ این روند، در واقع، بخشی از «معماری حکمرانی الگوریتمی» اسرائیل است که از سال‌ها پیش در سرزمین‌های اشغالی شکل گرفته است. این معماری به رژیم اجازه می‌دهد تا کنترل جمعیتی، نظارت امنیتی و سرکوب نظام‌مند را در قالب تصمیمات فنی و بی‌طرفانه جلوه دهد.

در چنین چارچوبی، الگوریتم‌ها به سپری برای مسئولیت‌گریزی سیاسی و حقوقی تبدیل می‌شوند. هنگامی که خطا یا تلفات گسترده رخ می‌دهد، تصمیم‌گیرندگان می‌توانند مسئولیت را به فناوری نسبت دهند مانند «اشتباه سامانه»، «سوگیری داده‌ها» یا «نقص در مدل یادگیری» و از پاسخگویی در برابر حقوق بین‌الملل بگریزند. در این وضعیت تصمیم انسانی در لایه‌های پیچیده‌ای از کد و داده پنهان می‌شود تا اراده‌ی سیاسی پشت حملات، نامرئی جلوه کند. این پدیده مصداقی از «فریب اخلاقی فناورانه» است و ظاهر بی‌طرف فناوری، ماهیت تبعیض‌آمیز و خشونت‌بار تصمیم را می‌پوشاند. در این معنا، ماشین نه تنها ابزار اجرای سیاست، بلکه بخشی از روایت‌سازی سیاسی خواهد بود.

ترجمه مقاله

استفاده از سامانه "لاوندر" در غزه و قانون انتخاب اهداف در عملیات نظامی:

سامانه‌های پشتیبان تصمیم‌گیری هوش‌مصنوعی و فناوری تشخیص چهره

امیلی اندرسین[۲]

پژوهشگر دکتری رشته مدیریت و روابط بین‌الملل ، دانشگاه لیدن، لاهه، هلند

چکیده

پس از ۷ اکتبر ۲۰۲۳، درگیری طولانی‌مدت بین اسرائیل و حماس به‌طور چشمگیری از نظر گستردگی و شدت خشونت افزایش یافت. گزارش‌ها نشان می‌دهند که ارتش اسرائیل از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) برای شناسایی افراد در موقعیت‌های هدف‌گیری استفاده کرده است. همچنین گزارش دیگری از به‌کارگیری فناوری تشخیص چهره (FRT) توسط ارتش اسرائیل برای شناسایی فلسطینیان در غزه خبر می‌دهد. پژوهشگران برطبق حقوق بین‌الملل بشردوستانه (IHL)، مشروعیت استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی بر هوش مصنوعی (AI-DSS) و میزان اتکای فرماندهان نظامی به هوش مصنوعی در زمینه اتخاذ تصمیم جهت هدف‌گیری افراد را زیرسوال می‌برند. این مقاله چالش‌های تعامل انسان و ماشین را با تمرکز بر توصیه‌های تولیدشده توسط الگوریتم‌ها و مسئولیت فرماندهان نظامی بررسی می‌کند. نتایج نشان می‌دهد که هرچند استفاده از سامانه‌های تشخیص چهره (FRT) می‌تواند دقت شناسایی افراد را افزایش دهد و پایبندی به تعهدات حقوق بین‌الملل بشردوستانه (IHL) را تقویت کند، کارآمدی آن به شرایط عملیاتی وابسته است. همچنین اهمیت ارتقای سواد فنی فرماندهان نظامی در زمینه سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) و اطمینان از صرف زمان کافی برای تأیید صحت اهداف تولیدشده توسط الگوریتم مورد تأکید قرار گرفته است.

[2] *Emelie Andersin*, PhD Fellow in Governance and Global Affairs

۱. مقدمه

جنگ مدرن به نبردی تکنولوژیک تبدیل شـــده اســـت. نیروهای نظامی باید با حجم عظیمی از اطلاعات، پیچیدگی عملیات و فشـــار زمانی شـــدید مقابله کنند. برای کارآمدی، ارتش‌ها نیاز دارند حجم زیادی از داده‌ها را به‌ویژه در محیط‌های پیچیده و پرتنش ارزیابی کنند.[۳] به همین دلیل، کشورها روزبه‌روز بیشتر به توسعه ابزارهای هوش مصنوعی (AI) گرایش پیدا کرده‌اند تا بتوانند این حجم عظیم اطلاعات را ســـریع‌تر پردازش کنند.[۴] برخی ارتش‌ها، مانند ایالات متحده، ســـامانه‌های مبتنی‌بر هوش مصـــنوعی طراحی کرده‌اند که با تولید توصـــیه‌هایی به فرماندهان نظامی در اتخاذ تصـــمیمات کمک می‌کنند.[۵] ســـامانه‌های پشـــتیبانی تصـــمیم‌گیری مبتنی‌بر هوش مصـــنوعی (AI-DSS) ابزارهایی رایانه‌ای هســـتند که برای کمک به انســـان‌ها در تصـــمیم‌گیری‌های پیچیده طراحی شـــده و اطلاعات مرتبط را ارائه می‌کنند.[۶] این ســـامانه‌ها به ارتش‌ها در جمع‌آوری داده‌ها، پیش‌بینی الگوها و ارائه توصیه‌های عملیاتی براساس حجم بالای اطلاعات کمک می‌کنند.[۷]

همزمان، استفاده از فناوری تشخیص چهره (FRT)[۸] در زمان جنگ نیز رو به افزایش است.[۹] کشورها از این فناوری بهره می‌برند زیرا ســـامانه‌های تشـــخیص چهره (FRT) می‌تواند به ســـامانه‌های هوش مصـــنوعی ورودی دهد تا اهداف انســـانی را براســـاس ویژگی‌های ظاهری شناسایی کنند.[۱۰] این فناوری می‌تواند دقت و اثربخشی در شناسایی یا تأیید هویت دشمنان شناخته‌شده از راه دور را افزایش دهد.[۱۱] به‌طور کلی، این ابزارهای هوش مصنوعی و تشخیص چهره برای تصمیم‌گیری سریع، مؤثر و دقیق در موقعیت‌های نظامی توسعه یافته‌اند.

در ۷ اکتبر ۲۰۲۳، درگیری بین اسرائیل و حماس پس از حمله حماس و دیگر گروه‌های مسلح به جنوب اسرائیل شدت گرفت. این حمله شـامل کشـتار، گروگان‌گیری و خشـونت جنسـی علیه غیرنظامیان بود[۱۲].[۱۳] در پاسـخ، اسـرائیل عملیات نظامی به نام «شـمشـیرهای آهنین» را در نوار غزه آغاز کرد. مجله +۹۷۲ و Local Call [۱۴] گزارش دادند که نیروهای دفاعی اسـرائیل (IDF)

---

[3] Merel A C Ekelhof, 'Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting' (2018) 71 Naval War College Review 61, 76.

[4] Ashley Deeks, 'Coding The Law of Armed Conflict: First Steps' in Matthew C Waxman and Thomas W Oakley (eds), *The Future of Armed Conflict* (Oxford University Press 2022), 45.

[5] Sydney Freedberg, 'atlas: Killer Robot? No. Virtual Crewman? Yes.' (*Breaking Defense*, 4 March 2019) <https://breakingdefense.com/2019/03/atlas-killer-robot-no-virtual-crewman-yes/>; Dustin Lewis, Naz Modirzadeh, and Gabriella Blum, 'The Pentagon's New Algorithmic-Warfare Team' (*Lawfare*, 26 June 2017) <https://www.lawfaremedia.org/article/pentagons-new-algorithmic-warfare-team>.

[6] Arthur Holland Michel, 'Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making' (May 2024) icrc Observations on External Report, 13.

[7] Nehal Bhuta, Susanne Beck, and Robin Geiβ, 'Present Futures: Concluding Reflections and Open Questions on Autonomous Weapons Systems' in Nehal Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016), 347–383.

[8] Anil K Jain, Arun Ross, and Salil Prabhakar, 'An Introduction to Biometric Recognition' (2004) 14 ieee Transactions on Circuits and Systems for Video Technology.

[9] Alison Mitchell, 'Distinguishing Friend from Foe: Law and Policy in the Age of Battlefield Biometrics' (2012) 50 Canadian Yearbook of International Law 289; William C Buhrow, *Biometrics in Support of Military Operations: Lessons from the Battlefield* (crs Press 2017).

[10] William H Boothby, 'Biometrics' in William H Boothby (ed), *New Technologies and the Law in War and Peace* (Cambridge University Press 2021), 397.

[11] Leah West, 'Face Value: Precaution versus Privacy in Armed Conflict' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (nato ccdcoe Publications 2022), 136.

[12] به دلیل رعایت امانت در ترجمه مجبور به بازگویی این تعابیر شدیم. هرچند خواننده باید بداند هیچ شواهدی مبنی‌بر خشونت جنسی از سوی مردم فلسطین علیه اتباع اسرائیل نیست و در اینجا نیز نویسنده صرفاً خواسته نوعی بی‌طرفی را رعایت کند تا مخاطبان بیشتری را جذب کند (مترجم).

[13] Abdelali Ragad et al, 'How Hamas Built a Force to Attack Israel on 7 October' (*bbc*, 27 November 2023) <https://www.bbc.com/news/world-middle-east-67480680>.

[14] The +972 Magazine and the Local Call are independent and non-profit magazines.

از سامانه پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) به نام لاوندر (Lavender) در غزه استفاده کرده‌اند.[۱۵] براساس مصاحبه با افسران اطلاعاتی IDF، این سامانه برای شناسایی فلسطینیانی طراحی شده است که ممکن است به حماس یا گروه‌های مسلح دیگر، مانند جهاد اسلامی فلسطین، وابستگی داشته باشند و آن‌ها را به‌عنوان اهداف بالقوه معرفی می‌کند. توصیه‌های لاوندر ابتدا به تحلیلگران اطلاعاتی ارسال می‌شود و پس از بررسی، گاهی نسخه بازبینی شده به فرماندهان نظامی فرستاده می‌شود.[۱۶] تصمیم نهایی برای تأیید حملات بر عهده فرمانده نظامی است.[۱۷] با این‌حال، دقیقاً مشخص نیست سامانه از چه منابعی برای ارائه توصیه استفاده می‌کند. به گفته IDF، این توصیه‌ها از داده‌های مختلفی مانند اطلاعات جغرافیایی-مکانی، اطلاعات سیگنالی، منابع انسانی و اطلاعات منبع باز حاصل می‌شوند.[۱۸]

در ۲۴ مارس ۲۰۲۴، نیویورک تایمز گزارش داد که ارتش اسرائیل یک برنامه نظارت انبوه مبتنی‌بر تشخیص چهره در غزه ایجاد کرده است که افراد را بدون رضایت یا اطلاع آن‌ها شناسایی می‌کند. این برنامه از دوربین‌ها، ایست‌های بازرسی نظامی و فیلم‌های پهپاد برای شناسایی افراد استفاده می‌کند.[۱۹] در سال ۲۰۲۳ یکی از فرماندهان یگان ۸۲۰۰ ارتش اسرائیل توضیح داد که آن‌ها می‌توانند افراد «خطرناک» را بر اساس فهرستی از اشخاص ثبت‌شده در سامانه شناسایی کنند.[۲۰] این موضوع نشان می‌دهد که ارتش اسرائیل احتمالاً از سامانه‌های تشخیص چهره (FRT) برای شناسایی افراد شناخته‌شده، بر پایه داده‌های زیست‌سنجی ذخیره‌شده در پایگاه داده، استفاده می‌کند.[۲۱]

طبق گزارش‌های مجله ۹۷۲+ و Local Call، سامانه «لاوندر» در شش هفته نخست درگیری، دست‌کم ۳۷ هزار هدف بالقوه شناسایی کرد که هم افراد رده‌بالا و هم نیروهای پایین‌رتبه را شامل می‌شد.[۲۲] گزارش‌ها همچنین حاکی از آن است که دقت این سامانه حدود ۹۰ درصد برآورد شده و ارتش اسرائیل گاهی حملات هوایی خود را براساس توصیه‌های این سامانه تأیید می‌کرد.

ارتش اسرائیل همچنین از سامانه ردیابی دیگری با نام « Where's Daddy? » استفاده کرده است؛ سامانه‌ای که افراد مظنون به فعالیت‌های نظامی را زیر نظر گرفته و هنگام ورود آن‌ها به خانه‌هایشان، ارتش را از این موضوع آگاه می‌سازد. در برخی موارد، خانه‌های این افراد برای بمباران علامت‌گذاری شده است، حتی زمانی‌که اعضای خانواده در خانه حضور داشتند.[۲۳]

ارتش اسرائیل از سامانه هوش مصنوعی دیگری به نام «Fire Factory» استفاده می‌کند. این سامانه با تکیه‌بر داده‌های مربوط به اهداف تأییدشده نظامی، میزان مهمات موردنیاز را محاسبه کرده، هزاران هدف را اولویت‌بندی می‌کند، آن‌ها را به هواپیماها

---

[15] Yuval Abraham, ‘ “Lavender”: The ai Machine Directing Israel’s Bombing Spree in Gaza’ (*+972 Magazine*, 3 April 2024) <https://www.972mag.com/lavender-ai-israeli-army-gaza/>.
[16] ibid.
[17] idf Website, ‘The idf’s Use of Data Technologies in Intelligence Processing’ (*Israel Defense Forces*, 18 June 2024) <https://www.idf.il/210062>.
[18] ibid.
[19] Sheera Frenkel, ‘Israel Deploys Expansive Facial Recognition Program in Gaza’ (*The New York Times*, 27 March 2024) <https://www.nytimes.com/2024/03/27/technology/israel-facial-recognition-gaza.html>.
[20] רורט ידעי לש רתוי ריהמ יוליגו גוויס תרשפאמ תיתוכאלמ הניב : ۸۲۰۰, תיתוכאלמה הניבה זכרמ ۱۷ דקפמ [Commander of the Artificial Intelligence Center, 8200: Artificial Intelligence Enables Faster Classification and Detection of Terrorist Targets] (*Israel Defense*, 14 February 2023) <https://www.israeldefense.co.il/node/57256>.
[21] For more details regarding previous collection, see Privacy International, ‘Biometrics and Counter-Terrorism: Case Study of Israel/Palestine’ (May 2021) <https://privacyinternational.org/sites/default/files/2021-06/PI%20Counterterrorism%20and%20Biometrics%20Report%20Israel_Palestine%20v7.pdf>, 9.
[22] Abraham (n 12).
[23] ibid.

و پهپادها اختصـــاص می‌دهد و حتی برنامه زمانی پیشـــنهادی برای اجرای حملات ارائه می‌دهد.[۲۴] به گفته تال میمران (Tal Mimran) و گال دهان (Gal Dahan) ، این سامانه برای تحلیل داده‌های اهداف پیشین و نیز «اولویت‌بندی و تخصیص اهداف» مورد استفاده قرار می‌گیرد.[۲۵]

لاوندر همچنین در کنار سـامانه پشـتیبانی تصـمیم‌گیری مبتنی‌بر هوش مصـنوعی (AI-DSS) دیگری به نام «the Gospel» استفاده می‌شود که ساختمان‌ها و سازه‌ها را به‌عنوان اهدافی علامت‌گذاری می‌کند که مظنونان از آن‌ها عملیات انجام می‌دهند. این سامانه‌ها توسط یگان اطلاعاتی نخبه ۸۲۰۰ ارتش اداره می‌شوند.[۲۶] با این‌حال، گزارش‌ها درباره استفاده از لاوندر محدود و تأیید آن‌ها دشوار است. با توجه به آسیب بی‌سابقه به غیرنظامیان و تخریب در غزه، پرداختن به این گزارش‌ها اهمیت بالایی دارد.

پژوهشگران و کارشناسان نگرانی عمیقی درباره آسیب گسترده به غیرنظامیان غزه ابراز کرده‌اند.[۲۷] تا زمان نوشتن این مقاله، بیش از ۴۰ هزار فلسطینی از ۷ اکتبر ۲۰۲۳ در غزه کشته شده‌اند[۲۸]، حداقل ۹۲.۴۰۱ نفر زخمی شده‌اند و بیش از نیمی از ساختمان‌های غزه تخریب یا آسیب دیده است.[۲۹] این سطح از تلفات و تخریب غیرنظامیان نگرانی جدی درباره استفاده از هوش مصنوعی در تصمیم‌گیری‌های هدفگیری نظامی و توانایی آن در کاهش آسیب غیرنظامیان ایجاد می‌کند.

هدف این مقاله بررسی چالش‌های حقوقی ناشی از اتکای فرماندهان نظامی به سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) در موقعیت‌های هدفگیری است. همچنین نقش تعامل انسان و ماشین، به ویژه استفاده فرماندهان از «اهداف» تولیدشده توسط الگوریتم‌ها، بررسی خواهد شد. لاوندر به‌عنوان مطالعه موردی مورد بررسی قرار می‌گیرد تا درک بهتری از چگونگی اعمال حقوق بین‌الملل بشردوستانه (IHL) بر استفاده از چنین سامانه‌هایی به دست آید.

این مقاله بر سامانه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) تمرکز دارد که ورودی خود را از سامانه‌های تشخیص چهره (FRT) دریافت می‌کنند، زیرا این حوزه کمتر در حقوق بین‌الملل بشردوستانه (IHL) بررسی شده است. بررسی این شکاف نشان می‌دهد که ترکیب سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) و سامانه‌های تشخیص چهره (FRT) ممکن است نگرانی‌های حقوقی قابل‌توجهی ایجاد کند. عدم دقت سامانه‌های تشخیص چهره (FRT) و توانایی سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) در پردازش حجم عظیم داده‌ها می‌تواند در درگیری‌های مسلحانه پیامدهای پیش‌بینی‌نشده‌ای داشته باشد. این مقاله سایر حوزه‌های حقوقی مرتبط، مانند

---

[24] Marissa Newman, ‘Israel Quietly Embeds ai Systems in Deadly Military Operations’ (*Bloomberg*, 16 July 2023) <https://www.bloomberg.com/news/articles/2023-07-16/israel-using-ai-systems-to-plan-deadly-militaryoperations?embedded-checkout=true&leadSource=uverify%20wall>.

[25] Tal Mimran and Gal Dahan, ‘Artificial Intelligence in the Battlefield: A Perspective from Israel’ (*Opinio Juris*, 20 April 2024) <https://opiniojuris.org/2024/04/20 /artificial-intelligence-in-the-battlefield-a-perspective-fromisrael/>.

[26] Yuval Abraham, ‘ ‘A Mass Assassination Factory’ : Inside Israel’ s Calculated Bombing of Gaza’ (*+972 Magazine*, 30 November 2023) <https://www.972mag.com/mass -assassination-factory-israel-calculated-bombing-gaza/>.

[27] University Network for Human Rights et al, ‘Genocide in Gaza: Analysis of International Law and Its Application to Israel’ s Military Actions Since October 7, 2023’ (15 May 2024) <https://static1.squarespace.com/static/66a134337e960f229da81434/t/66fb05bb0497da4726e125d8/1727727037094/Genocide+in+Gaza+-+Final+version+051524.pdf>.

[28] تا زمان ترجمه این مقاله حداقل ۶۸ هزار و ۵۲۷ نفر شهید شده‌اند

[29] Julia Frankel, ‘With Gaza’ s Death Toll over 40,000, Here’ s the Conflict by Numbers’ (*The Associated Press*, 15 August 2024) <https://apnews.com/article /israel-hamas-gaza-war-palestinians-statistics 400007ebec13101f6d08fe10cedbf5e172dde>.

حقوق بین‌الملل بشردوستانه زنده (IHRL) [۳۰] و قوانین محافظت از داده‌های شخصی (DPL) [۳۱] را بررسی نمی‌کند، اما این بدان معنا نیست که اهمیت این حوزه‌ها نادیده گرفته شده است، بلکه خارج از محدوده این مطالعه قرار دارند.

۲. استفاده از فناوری در درگیری‌های مسلحانه

این بخش به بررسی فناوری‌هایی می‌پردازد که برای شناسایی هویت افراد از طریق سامانه‌های تشخیص چهره (FRT) و همچنین تولید توصیه‌هایی برای هدف‌گیری استفاده می‌شوند. در ابتدا، به‌طور خلاصه توضیح داده می‌شود که سامانه‌های تشخیص چهره (FRT) چیستند و چگونه در حوزه‌های غیرنظامی و در شرایط درگیری مسلحانه به‌کار گرفته می‌شوند. سپس، سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) ، شیوه عملکرد و کاربردهای نظامی آن‌ها مورد بحث قرار می‌گیرد.

۲.۱. استفاده از سامانه‌های تشخیص چهره (FRT) برای شناسایی یا تأیید هویت افراد

هر انسان به‌واسطه ویژگی‌های چهره‌اش دارای خصوصیات منحصر به‌فردی است.[۳۲] تشخیص چهره یک روش زیست‌سنجی[۳۳] است که با شناسایی خودکار ویژگی‌های صورت، سعی می‌کند هویت یک فرد را تشخیص دهد یا تأیید کند.[۳۴] سامانه‌های تشخیص چهره (FRT) معمولاً مبتنی‌بر هوش مصنوعی هستند و از الگوریتم‌ها و روش‌های یادگیری ماشین برای شناسایی، پردازش و تطبیق استفاده می‌کنند.[۳۵] این فرآیند، چهره را به‌عنوان شاخصی برای هویت در نظر می‌گیرد:[۳۶] سامانه ابتدا اطلاعات چهره را ثبت یا جمع‌آوری کرده، الگوهای چهره را استخراج و آن‌ها را به بازنمایی‌های ریاضی تبدیل می‌کند و سپس این بازنمایی‌ها را با داده‌های زیست‌سنجی ذخیره‌شده در یک پایگاه داده مقایسه می‌کند تا هویت فرد را بیابد.[۳۷] هدف‌های اصلی تشخیص چهره سه‌گانه‌اند: (۱) شناسایی، (۲) تأیید و (۳) طبقه‌بندی هویت فرد.

در فرآیند شناسایی، سامانه زیست‌سنجی نمونه‌ی جدید را در مقابل تمام داده‌های قبلی قرار داده و از تطبیق یک-به-چند برای یافتن شخص ناشناس استفاده می‌کند. در فرآیند تأیید، سامانه یک مقایسه یک-به-یک انجام می‌دهد تا مشخص شود آیا

---

[30] West (n 9).
[31] Asaf Lubin, 'The Rights to Privacy and Data Protection under International Humanitarian Law and Human Rights Law' in Robert Kolb, Gloria Gaggioli, and Pavle Kilibarda (eds), *Research Handbook on Human Rights and Humanitarian Law: Further Reflections and Perspectives* (Edward Elgar Publishing 2022).
[32] Marcus Smith and Seumas Miller, *Biometric Identification, Law and Ethics* (Springbriefs in Ethics 2021), 22.
[33] Biometrics Institute, 'Types of Biometrics' <https://www.biometricsinstitute.org/what-is-biometrics/types-ofbiometrics/> accessed 17 January 2024.
[34] Jain, Ross and Prabhakar (n 6).
[35] Smith and Miller (n 28), 22-23.
[36] Kelly A Gates, *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance* (New York University Press 2011), 8.
[37] For more details, see Gates (n 32), 8; Lucas Introna and Helen Nissenbaum, 'Facial Recognition Technology: A Survey of Policy and Implementation Issues' (2009) Center for Catastrophe Preparedness and Response, 15-16.

فردی که ادعای هویت کرده همان شخص مورد ادعاست یا خیر.[۳۸] ر فرآیند طبقه‌بندی، سامانه با استخراج ویژگی‌های مختلف از چهره، اطلاعاتی مانند حالت احساسی،[۳۹]، جنسیت یا گروه نژادی فرد[۴۰] را شناسایی و دسته‌بندی می‌کند.

خروجی سامانه‌های تشخیص چهره (FRT) معمولاً به‌صورت پاسخ قطعی «بله» یا «خیر» نیست، بلکه نتیجه را به‌صورت یک احتمال ارائه می‌دهد. این الگوریتم‌ها معمولاً «آستانه اطمینان» تولید می‌کنند که نشان می‌دهد دو تصویر چهره تا چه اندازه ممکن است متعلق به یک فرد باشند.[۴۱] هرچه این امتیاز بالاتر باشد، احتمال تطابق درست میان دو چهره بیشتر است.[۴۲]

در این سامانه‌ها معمولاً دو نوع خطا وجود دارد: (۱) مثبت کاذب و (۲) منفی کاذب. خطای مثبت کاذب زمانی رخ می‌دهد که سامانه به اشتباه اعلام کند چهره فرد با یکی از نمونه‌های موجود در پایگاه داده مطابقت دارد، در حالی‌که این تطبیق نادرست است. در مقابل، خطای منفی کاذب زمانی اتفاق می‌افتد که سامانه نتواند چهره فرد را با هیچ تصویری در پایگاه داده تطبیق دهد، با اینکه آن فرد در واقع در پایگاه داده ثبت شده است..[۴۳]

آستانه‌ای که برای پذیرش یا رد یک تطابق تعیین می‌شود، به چند عامل بستگی دارد: کنترل‌شده بودن محیط، حضور ناظر انسانی برای نظارت‌بر سامانه و میزان حساسیت محیط موردنظر. اگر آستانه اطمینان بالا تنظیم شود تا از وقوع مثبت‌های کاذب جلوگیری شود، احتمال ایجاد منفی‌های کاذب بیشتر خواهد شد؛ برعکس، اگر آستانه پایین‌تر باشد، احتمال مثبت‌های کاذب افزایش پیدا می‌کند.[۴۴]

این فناوری روزبه‌روز نقش بیشتری در حوزه امنیت پیدا کرده است. سامانه‌های تشخیص چهره (FRT) می‌توانند مظنونان را در اماکن عمومی شناسایی کنند و معمولاً توسط نهادهای مجری قانون به‌کار گرفته می‌شوند. این فناوری‌ها همچنین می‌توانند برای مقابله با تروریسم استفاده شوند؛ برای مثال، تصاویر ضبط‌شده توسط دوربین‌های مدار بسته (CCTV) را با پایگاه داده‌های چهره افراد تحت نظر مقایسه می‌کنند.[۴۵] به‌عنوان نمونه، مقامات اسرائیلی از سامانه‌های تشخیص چهره (FRT) برای نظارت‌بر غیرنظامیان ساکن در سرزمین‌های فلسطینی اشغالی استفاده می‌کنند.[۴۶] همزمان، کاربرد تشخیص چهره در درگیری‌های

---

[38] Mitchell (n 7).
[39] Joy Buolamwini et al, 'Facial Recognition Technologies: A Primer' (29 May 2020) <https ://globaluploads.webflow.com/5e027ca188c99e3515b404b7/5ed1002058516c11edc66a14 _FRTsPrimerMay2020.pdf>.
[40] European Union Agency for Fundamental Rights, 'Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement' (2019) - ≤https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition - technology-focus-paper.pdf> accessed 7 January 2024.
[41] ibid, 9.
[42] European Data Protection Board, 'Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement' (Version 2.0, adopted on 26 April 2023).
[43] William Crumpler, 'How Accurate are Facial Recognition Systems – and Why Does It Matter?' (*Center for Strategic & International Studies*, 14 April 2020) <https://www.csis.org/blogs/strategic-technologies-blog/how-accurateare-facial - recognition-systems-and-why-does-it>.
[44] Ibid 39.
[45] Information Commissioner's Opinion, 'The Use of Live Facial Recognition Technology in Public Places' (18 June 2021) < https://ico.org.uk/media/2619985/ico-opinion-the-use-of -lfr-in-public-places-20210618.pdf>.
[46] For more extensive details, see Rohan Talbot, 'Automating Occupation: International Humanitarian and Human Rights Law Implications of the Deployment of Facial Recognition Technologies in the Occupied Palestinian Territory' (2020) 102 International Review of the Red Cross 823; Omar Yousef Shehabi, 'Emerging Technologies, Digital Privacy, and Data Protection in Military Occupation' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (nato ccdcoe Publications 2022).

مسلحانه نیز گسترده‌تر شده است.[۴۷] ایالات متحده پس از حمله سال ۲۰۰۳ به عراق، پیشگام استفاده از زیست‌سنجی در عملیات نظامی شد[۴۸]؛ بین سال‌های ۲۰۰۸ تا ۲۰۱۷، این فناوری برای دستگیری یا هدف‌گیری ۱۷۰۰ نفر به کار گرفته شد و دسترسی ۹۲.۰۰۰ نفر را به تأسیسات نظامی محدود کرد.[۴۹] همچنین، پس از حمله ۷ اکتبر، مقامات اسرائیلی برای شناسایی کشته‌شدگان به سامانه‌های تشخیص چهره (FRT) متکی بودند.[۵۰] ارتش اوکراین نیز در جنگ با روسیه از این فناوری برای شناسایی کشته‌شدگان خود و نیروهای روس استفاده کرد.[۵۱]

۲.۲. استفاده از AI-DSS برای یافتن، انتخاب و ارائه توصیه هدف‌ها

سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) به‌طور کلی یا برپایه مدل‌های قطعی (deterministic) هستند یا مدل‌های غیرقطعی (non-deterministic).[۵۲] سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی سنتی بر پایه مدل‌های قطعی ساخته می‌شوند، یعنی همیشه خروجی یکسان و قابل پیش‌بینی تولید می‌کنند و هیچ تصادفی در کارشان وجود ندارد.[۵۳] این سامانه‌ها هرچند قابل اعتمادند، اما تولیایی محدودی در پردازش مسائل پیچیده و واقعیت‌های غیرقابل پیش‌بینی دارند.[۵۴]

در مقابل، مدل‌های غیرقطعی (یا تصادفی) نتایجی ارائه می‌کنند که شامل سطحی از عدم قطعیت و غیرقابل پیش‌بینی بودن است. این مدل‌ها با داده‌های آموزشی تغذیه می‌شوند؛ ورودی‌ها همراه با نمونه‌هایی از خروجی مطلوب در اختیار الگوریتم قرار می‌گیرند تا سامانه یاد بگیرد.[۵۵] بنابراین، آن‌ها محدود به مقادیر کدگذاری‌شده نیستند و می‌توانند واقعیت‌های پیچیده‌تر را مدل‌سازی کنند. این ویژگی در حوزه نظامی بسیار مفید است، زیرا میدان نبرد همواره غیرقابل پیش‌بینی است و پیچیدگی‌های دنیای واقعی را باید در نظر گرفت. با توجه به عدم قطعیت‌ها و فقدان دانش کامل، ممکن است نتوان دقیقاً توضیح داد که چرا مدل یک خروجی خاص تولید کرده است.[۵۶]

---

[47] Marten Zwanenburg, 'Know Thy Enemy: The Use of Biometrics in Military Operations and International Humanitarian Law' (2021) 97 International Law Studies 1404.

[48] Spencer Ackerman, 'U.S. Holds On to Biometrics Database of 3 Million Iraqis' (*Wired*, 21 December 2011) <https://www.wired.com/2011/12/iraq-biometrics-database/>; Annie Jacobsen, *First Platoon: A Story of Modern War in the Age of Identity Dominance* (Penguin 2021).

[49] United States Government Accountability Office, 'dod Biometrics and Forensics: Progress Made in Establishing Long-Term Deployable Capabilities, but Further Actions are Needed' (August 2017) <https://www.gao.gov/assets/gao-17-580.pdf> accessed 2 February 2024.

[50] Masha Borak, 'Israel is Using Amazon Rekognition to Locate Missing and Dead' (*Biometric Update*, 23 October 2023) <https://www.biometricupdate.com/202310 /israel-is-using-amazon-rekognition-to-locate-missing-anddead>.

[51] Drew Harwell, 'Ukraine is Scanning Faces of Dead Russians, then Contacting the Mothers' (*Washington Post*, 15 April 2022) <https://www.washingtonpost.com /technology/2022/04/15/ukraine-facial-recognition-warfare/>.

[52] Holland Michel (n 4) ,18.

[53] Agnieszka Lazarowska, 'A New Deterministic Approach in a Decision Support System for Ship's Trajectory Planning ' (2017) 71 Expert Systems with Applications 469; Jorge Vargas Florez et al, 'A Decision Support System for Robust Humanitarian Facility Location' (2015) 46 Engineering Applications of Artificial Intelligence 326.

[54] Priya Narayanan et al, 'First Year Report of arl Director's Strategic Initiative (fy20-23) : Artificial Intelligence (ai) for Command and Control (C2) of Multi-Domain Operations (mdo) ' (devcom Army Research Laboratory, May 2021), 2.

[55] Holland Michel (n 4), 18.

[56] Thomas W Lucas, 'The Stochastic Versus Deterministic Argument for Combat Simulations: Tales of When the Average Won't Do' (2000) 5 Military Operations Research 9; Timothy J Horrigan, 'Configuration and the Effectiveness of Air Defense Systems in Simplified, Idealized Combat Situations - A Preliminary Examination' (Horrigan Analytics, June 1995), 5.

با کمک یادگیری ماشین، برنامه‌نویس سامانه را آموزش می‌دهد تا وظایف الگوریتم را انجام دهد و در عین حال تجربه‌ای در ارائه توصیه‌ها کسب کند.[۵۷] سامانه الگوها و ویژگی‌های موجود در داده‌ها را شناسایی کرده و براساس داده‌های ورودی، خروجی تولید می‌کند.[۵۸] این فرآیند سرعت تصمیم‌گیری را افزایش داده و امکان شناسایی الگوها در حجم بالای داده‌ها را فراهم می‌کند.[۵۹] الگوریتم، مجموعه‌ای از مراحل محاسباتی است که ورودی را به خروجی تبدیل می‌کند و می‌تولند به دو روش آموزش ببینند:[۶۰]

- **یادگیری تحت نظارت:** الگوریتم با داده‌های برچسب‌خورده آموزش داده می‌شود و بازخورد برای طبقه‌بندی یا رگرسیون دریافت می‌کند.
- **یادگیری بدون نظارت:** الگوریتم بدون نظارت، خود به خود الگوها را کشف می‌کند و داده‌های بدون برچسب را خوشه‌بندی می‌کند.[۶۱]

برای مثال، اگر فردی ویژگی‌هایی مشابه کسی که به‌عنوان غیرنظامی درگیر مستقیم در درگیری‌ها (DPH) شناسایی شده است داشته باشد، سامانه ممکن است آن فرد را نیز به‌عنوان DPH برچسب‌گذاری کند. از آنجا که مدل‌های غیرقطعی مبتنی‌بر احتمال هستند، نتایج آن‌ها همیشه قطعی نیست و براساس احتمال وقوع وضعیت DPH ارائه می‌شود، نه براساس قطعیت مطلق.

دقت یک سامانه هوش مصنوعی نشان می‌دهد که این سامانه تا چه اندازه به‌طور مداوم پیش‌بینی‌های درست انجام می‌دهد و معمولاً میزان خطای کمی دارد. دقت معمولاً به صورت عددی بین ۰ تا ۱ یا ۰ تا ۱۰۰ گزارش می‌شود؛ جایی که ۰ به معنای همیشه نادرست بودن پیش‌بینی و ۱ یا ۱۰۰ به معنای همیشه درست بودن آن است. بنابراین، مدل‌های غیرقطعی براساس برآوردهای احتمالی عمل می‌کنند و همیشه یک حاشیه خطا دارند.

دقت مدل معمولاً با استفاده از ماتریس درهم‌ریختگی (confusion matrix) سنجیده می‌شود. این ماتریس عملکرد پیش‌بینی مدل و خطاهای آن را نشان می‌دهد[۶۲] و شامل چهار دسته است:

۱. مثبت‌های حقیقی (True Positives) : تطابق درست واقعی
۲. منفی‌های حقیقی (True Negatives) : تطابق نادرست واقعی
۳. مثبت‌های کاذب (خطای نوع I) [۶۳]
۴. منفی‌های کاذب (خطای نوع II) [۶۴]

---

[57] Katrina Wakefield, 'Predictive Modeling Analytics and Machine Learning' (sas Data and ai Solutions) <https://www.sas.com/en_gb/insights/articles/analytics/a-guide-to-predictive-analytics -and-machinelearning.html> accessed 10 February 2024.

[58] The Pecan Team, 'Contrasting Generative ai, Predictive ai, and Machine Learning' (*Pecan*, 6 December 2023) <https://www.pecan.ai/blog/generative -ai-predictive-ai-machine-learning/>.

[59] Avi Goldfarb and Jon R Lindsay, 'Prediction and Judgement: Why Artificial Intelligence Increases the Importance of Humans in War' (2022) 46 International Security 7.

[60] Thomas H Cormen et al, *Introduction to Algorithms* (4th edn, mit Press 2022), 5.

[61] Pratap Dangeti, *Statistics for Machine Learning: Build Supervised, Unsupervised, and Reinforcement Learning Models Using Both Python and R* (Packt Publishing 2017), 8.

[62] Aniruddha Bhandari, 'Understanding & Interpreting Confusion Matrix in Machine Learning (Updated 2024)' (*Analytics Vidhya*, 11 January 2024) <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machinelearning/>.

[63] ibid. False positives or Type I errors occur when the value was falsely predicted, such that the actual value was negative but the model predicted a positive value.

[64] ibid. False negatives or Type ii errors occur when the predicted value was falsely predicted, such that the actual value was positive but the model predicted a negative error.

برای مثال، اگر الگوریتمی آموزش داده شده باشد تا اهداف قانونی (lawful) را شناسایی کند، مثبت‌های حقیقی نشان‌دهنده اهداف قانونی و منفی‌های حقیقی نشان‌دهنده اهداف غیرقانونی (unlawful) هستند. برای عملکرد مؤثر، الگوریتم باید استحکام (robustness) داشته باشد[۶۵] تا کارایی خود را حفظ کند و در برابر حملات متخاصم آسیب‌پذیر نباشد.[۶۶] همچنین، قابلیت اطمینان (reliability) سامانه با میزان اعتمادپذیری و تعداد خطاها و اثرات ناخواسته سنجیده می‌شود و نشان می‌دهد که سامانه تا چه حد عملکردی سازگار با هدف موردنظر دارد.[۶۷]

هوش مصنوعی به‌طور فزاینده‌ای در کاربردهای نظامی – که به آن «الگوریتم‌های جنگ» نیز گفته می‌شود[۶۸] – به ابزاری مهم در میدان نبرد تبدیل شده است. برای مثال، تیم الگوریتم-جنگ وزارت دفاع ایالات متحده (DoD) از فیلم‌های پهپادها در عراق و سوریه برای شناسایی اشیاء و برچسب‌گذاری داده‌ها استفاده کرد.[۶۹] به گفته تیم الگوریتم-جنگ وزارت دفاع ایالات متحده DoD، ابزارهای سنتی برای کاهش خسارت جانبی (collateral damage) نمی‌توانند همیشه پویایی‌های محیط عملیاتی را در مقایسه با ابزارهای غیرقطعی در نظر بگیرند.[۷۰]

۳. چالش‌های استفاده از سامانه‌های تشخیص چهره (FRT) و هوش مصنوعی در درگیری‌های مسلحانه

این بخش به بررسی چالش‌های استفاده از سامانه‌های تشخیص چهره (FRT) و هوش مصنوعی در عملیات نظامی می‌پردازد و آن‌ها را در چهار دسته اصلی بررسی می‌کند:

الف) دقت یا عدم دقت سامانه‌های تشخیص چهره

ب) سوگیری اتوماسیون

ج) تأثیر سوگیری‌های فنی و شناختی

د) ابهام یا عدم شفافیت هوش مصنوعی

---

[65] Ronan Hamon, Henrik Junklewitz, and Ignacio Sanchez, 'Robustness and Explainability of Artificial Intelligence' (*Publications Office of the European Union*, 2020).

[66] To read more about adversarial attacks, see Mark A Visger, 'Garbage in, Garbage Out: Data Poisoning Attacks and Their Legal Implications' in Laura A Dickinson and Edward W Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023).

[67] International Committee of the Red Cross, 'Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control' (Geneva, August 2019), 10.

[68] Dustin A Lewis, Gabriella Blum, and Naz K Modirzadeh, 'War-Algorithm Accountability' (Harvard Law School Program on International Law and Armed Conflict, 31 August 2016).

[69] Lewis, Modirzadeh, and Blum (n 3).

[70] US DoD, 'No-Strike and the Collateral Damage Estimation Methodology', (Chairman of the Joint Chiefs of Staff Instruction, cjcsi 3160.01, 12 October 2012).

۳.۱. (عدم) دقت سامانه‌های تشخیص چهره

تشخیص چهره یکی از دشوارترین روش‌های زیست‌سنجی است، زیرا چهره‌ها پیچیده و چندبعدی هستند و ثابت نمی‌مانند.[۷۱] این روش به‌دلیل تغییرات طبیعی ناشی از پیر شدن[۷۲]، آرایش[۷۳] یا معلولیت‌ها[۷۴]، یکی از کم‌دقت‌ترین روش‌های زیست‌سنجی شناخته می‌شود. در مقایسه، سایر روش‌های زیست‌سنجی مانند اسکن عنبیه یا شبکیه دقیق‌ترند، زیرا جمع‌آوری داده‌ها با تماس نزدیک انجام می‌شود، در حالی‌که تشخیص چهره از راه دور صورت می‌گیرد.[۷۵]

سامانه‌های تشخیص چهره (FRT) می‌توانند در محیط‌های مختلف استفاده شوند. در محیط‌های کنترل‌شده، عواملی مانند زاویه و نور بهتر قابل کنترل هستند؛ مانند کنترل گذرنامه.[۷۶] اما در محیط‌های کنترل‌نشده، این عوامل به‌طور مداوم قابل کنترل نیستند. افراد ممکن است حرکت کنند و در فضاهای عمومی با نور نامناسب قرار گیرند. تحقیقات نشان داده‌اند که دقت سامانه‌های تشخیص چهره (FRT) در محیط‌های کنترل‌نشده به‌طور قابل‌توجهی کاهش می‌یابد، به‌ویژه وقتی فرد مستقیماً به دوربین نگاه نمی‌کند.[۷۷] عامل دیگری که دقت را تحت تأثیر قرار می‌دهد، این است که تصاویر زنده هستند یا از قبل ضبط شده‌اند. فناوری تشخیص چهره (FRT) زنده (FRTL) ، چهره‌ها را از فیلم‌های ویدیویی استخراج می‌کند تا بررسی کند آیا فردی در پایگاه داده تصاویر وجود دارد یا خیر.[۷۸] سامانه‌های تشخیص چهره (FRT) زنده به دلیل ناتوانی در کنترل عواملی مانند فاصله، زاویه و نور معمولاً نرخ مثبت کاذب بالاتری دارد.[۷۹]

در حوزه غیرنظامی، نمونه‌هایی از مثبت‌های کاذب مشاهده شده است؛ یعنی زمانی که نیروهای اجرای قانون از سامانه‌های تشخیص چهره (FRT) برای شناسایی مظنونان استفاده کرده‌اند. به‌عنوان مثال، گزارشی درباره استفاده از این سامانه‌ها در بریتانیا نشان داد که به‌طور متوسط، نرخ شناسایی نادرست در سراسر کشور ۹۵٪ بوده است.[۸۰] همین مشکل در ایالات متحده هم رخ داده است؛ برای نمونه، در سال ۲۰۱۹، پلیس نیوجرسی یک مظنون را به اشتباه با فرد دیگری اشتباه گرفت[۸۱] و او به مدت ده روز در زندان ماند.[۸۲]

---

[71] Mary Clark, 'Top Five Biometrics (Face, Fingerprint, Iris, Palm and Voice) Modalities Comparison' <https://www.bayometric.com/biometrics-face-finger-iris-palm-voice/> accessed 17 February 2024.
[72] Leila Boussaad and Aldjia Boucetta, 'Deep-Learning Based Descriptions in Application to Aging Problem in Face Recognition' (2022) 34 Journal of King Saud University - Computer and Information Sciences 2975.
[73] Sayako Ueda and Takamasa Koyama, 'Influence of Make-up on Facial Recognition' (2010) 39 Perception 260.
[74] European Union Agency for Fundamental Rights (n 36).
[75] John D Woodward et al, *Army Biometric Applications: Identifying and Addressing Sociocultural Concerns* (rand 2001), 19.
[76] David Bolt, 'An Inspection of the Policies and Practices of the Home Office's Borders, Immigration and Citizenship Systems Relating to Charging and Fees' (Independent Chief Inspector, 2019).
[77] Patrick Grother, Mei Ngan and Kayee Hanaoka, 'Facial Recognition Technology Evaluation (frte) : Part 2: Identification ' (National Institute of Standards and Technology, September 2023) <https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf> [30 January 2025].
[78] European Union Agency for Fundamental Rights (n 36).
[79] Gates (n 32), 71.
[80] Big Brother Watch, 'Face Off: The Lawless Growth of Facial Recognition in UK Policing' (May 2018).
[81] Kashmir Hill, 'Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match' (*The New York Times,* 29 December 2020) <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>.
[82] Many other cases of misidentification have been reported: see Kashmir Hill, 'Wrongfully Accused by an Algorithm' (*The New York Times*, 24 June 2020) <https://www.nytimes.com/2020/06/24/technology/facialrecognition-arrest.html>; Kashmir Hill, 'Eight Months Pregnant and Arrested After False Facial Recognition Match' (*The New York Times*, 6 August 2023) <https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest
.html>; Johana Bhuiyan, 'Facial Recognition Used After Sunglass Hut Robbery Led to Man's Wrongful Jailing, says sui (*The Guardian*, 23 January 2024) <https://www.theguardian.com/technology/2024/jan/22/sunglass-hut-facialrecognition-wrongful-arrest-lawsuit>.

در محیط‌های درگیری مسلحانه، حضور غیرنظامیان، گرد و غبار و نور ناکافی باعث می‌شود تشخیص چهره برای سامانه‌های تشخیص چهره (FRT) دشوار شود. اگر سامانه‌ها بر روی داده‌هایی آموزش ندیده باشند که تنوع چهره‌ها در مناطق درگیری را نشان دهد، نرخ خطا برای گروه‌های کم‌نمایان بالاتر می‌رود. به‌ویژه، مثبت‌های کاذب خطرناک هستند، زیرا یک غیرنظامی ممکن است به اشتباه با یک عضو گروه مسلح تطبیق داده شود. شناسایی نادرست در این شرایط می‌تواند پیامدهای جدی‌تری نسبت به محیط‌های معمول اجرای قانون داشته باشد؛ بنابراین، استانداردهای سختگیرانه‌ای برای کاهش مثبت‌های کاذب در درگیری‌های مسلحانه ضروری است. حتی با نرخ پایین مثبت‌های کاذب، پیامدها می‌تواند فاجعه‌بار باشد؛ برای مثال، در یک جمعیت ۲۰۰,۰۰۰ نفری با نرخ مثبت کاذب ۰.۱٪، حدود ۲۰۰۰ نفر به اشتباه به عنوان هدف شناسایی می‌شوند. بنابراین، ارزیابی دقت باید براساس نرخ خطا، اندازه جمعیت و حساسیت محیط انجام شود.

۳.۲. سوگیری اتوماسیون

پژوهشگران معتقدند یکی از مزایای سامانه‌های مبتنی‌بر هوش مصنوعی این است که می‌توانند خطاهای انسانی را حذف کنند و احتمال اشتباه را کاهش دهند. از این رو، برخی بر این باورند که ماشین‌ها قابل اعتمادتر از انسان‌ها هستند، زیرا از واکنش‌های احساسی تأثیر نمی‌گیرند و قابلیت‌های انسانی را پشت سر می‌گذارند.[۸۳]

با این‌حال، انسان‌ها اغلب به‌طور افراطی به سامانه‌های پشتیبان تصمیم مبتنی‌بر رایانه اعتماد می‌کنند و اطلاعات متناقض را نادیده می‌گیرند یا جستجو نمی‌کنند.[۸۴] نمونه‌هایی از این اعتماد بیش از حد در حوزه‌های مختلف دیده شده است: در حوزه بهداشت و درمان، پزشکان به تشخیص‌های نادرست رسیدند[۸۵]؛ رانندگان وسایل نقلیه، پل تخریب‌شده را نادیده گرفتند و تلفات جانی رخ داد[۸۶]؛ دانش‌آموزان دستورالعمل‌های یک ربات را دنبال کردند و به یک ساختمان در حال سوختن رفتند.[۸۷] توصیه‌های الگوریتمی AI-DSS می‌تواند باعث سوگیری اتوماسیون شود؛ یعنی انسان‌ها توصیه‌های خودکار را به جای اطلاعات متناقضی که ممکن است خلاف آن‌ها باشد، ترجیح دهند.[۸۸] تحقیقات نشان داده‌اند که خطر سوگیری اتوماسیون در موقعیت‌های حساس به زمان افزایش می‌یابد.[۸۹]

مسئله دوم: سرعت تولید توصیه‌ها توسط سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS)

سرعت تولید توصیه‌ها توسط سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) بسیار بالاتر از توانایی انسانی است.[۹۰] فرماندهان نظامی ممکن است به دلیل دریافت پیوسته اطلاعات در زمان واقعی، تمایل پیدا کنند سریع‌تر عمل

---

[83] Robin Geiß and Henning Lahmann, 'Autonomous Weapons Systems: A Paradigm Shift for the Law of Armed Conflict?' in Jens David Ohlin (ed), *Research Handbook on Remote Warfare* (Edward Elgar Publishing 2017), 373.

[84] Elke Schwarz, 'Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control' (2021) 1 The Philosophical Journal of Conflict and Violence 53.

[85] Thomas Dratsch et al, 'Automation Bias in Mammography: The Impact of Artificial Intelligence bi-rads Suggestions on Reader Performance' (2023) 307 Radiology 1

[86] Jenny Gross, 'He Drove Into a Creek and Died. His Family Blames Google Maps' (*The New York Times*, 21 September 2023) <https://www.nytimes.com/2023/09/21/us/google-maps-lawsuit-collapsed-bridge.html>.

[87] John Toon, 'In Emergencies, Should You Trust a Robot?' (*Georgia Tech*, 29 February 2016) <https://news.gatech.edu/news/2016/02/29/emergencies-should-you-trust-robot>.

[88] Linda J Skitka, Kathleen Mosier, and Mark D Burdick, 'Accountability and Automation Bias' (2000) 52 International Journal of Human-Computer Studies 701.

[89] Mary L Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems' (2004) American Institute of Aeronautics and Astronautics.

[90] Berenice Boutin, 'Legal Questions to the Use of Autonomous Weapons Systems' (Briefing Paper to the aiv/cavv Advisory Report on Autonomous Weapon Systems: The Importance of Regulation and Investment, 2021), 4.

کرده و به توصیه‌های سامانه اعتماد کنند، زیرا احساس «فوریت» در آن‌ها افزایش می‌یابد. هرچه سامانه پیچیده‌تر و خودکارتر باشد، احتمال اعتماد فرماندهان به خروجی آن بیشتر می‌شود، زیرا آن‌ها حس می‌کنند زمان کافی برای تأیید هدف را در اختیار ندارند. این موضوع می‌تواند بر میزان حفظ قضاوت انسانی فرماندهان هنگام اتکا به توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS)، به ویژه در اجرای اقدامات احتیاطی مطابق ماده ۵۷ پروتکل الحاقی ۱، تأثیر بگذارد.[۹۱] در مناطق پرجمعیت و با حضور غیرنظامیان، اهمیت دارد که تصمیم‌گیرندگان زمان کافی برای ارزیابی و کاهش آسیب به غیرنظامیان صرف کنند.

۳.۳. تأثیر سوگیری‌های فنی و شناختی

نگرانی‌هایی وجود دارد که سامانه‌های هوش مصنوعی ممکن است تحت تأثیر سوگیری‌ها قرار گیرند و سوگیری‌های موجود در داده یا الگوریتم‌ها تقویت شوند، چه در مرحله آموزش و توسعه و چه در حین استفاده عملی.[۹۲] پژوهش‌ها نشان داده‌اند که انواع مختلفی از سوگیری‌های فنی در داده‌ها وجود دارد[۹۳] که می‌توانند بر توصیه‌های تولیدشده توسط هوش مصنوعی تأثیر بگذارند.[۹۴]

- **سوگیری الگوریتمی:** خطاهای سیستماتیک و مکرر در خروجی ناشی از داده‌های غیرنماینده در مرحله آموزش است. این سوگیری باعث می‌شود الگوریتم در شناسایی برخی گروه‌ها دقیق‌تر و در شناسایی سایر گروه‌ها کم‌دقت باشد.[۹۵]
- **سوگیری نمونه‌گیری:** زمانی رخ می‌دهد که برخی گروه‌ها از جمعیت بیش از دیگران احتمال انتخاب شدن داشته باشند.[۹۶]
- **سوگیری انتساب گروهی:** شامل کلیشه‌سازی نسبت به «گروه بیرونی» (افرادی که به یک گروه خاص تعلق ندارند) و تمایل به ترجیح دادن «گروه درونی» (گروهی خاص) است.[۹۷]
- **سوگیری عمل:** انسان‌ها اغلب عمل را به بی‌عملی ترجیح می‌دهند، حتی زمانی که شواهد کافی وجود ندارد یا اطلاعات کامل ندارند. این باعث می‌شود تصمیم‌گیری تحت فشار یا اجبار انجام شود.[۹۸]

---

[91] Protocol Additional (i) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims in International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 unts 3 (Protocol) art 57 (‘ap i’).

[92] icrc, ‘Artificial Intelligence and Machine Learning in Armed Conflict: A Human Centred Approach’ (2020) 102 International Review of the Red Cross 463.

[93] Lindsey Jacques, ‘Facial Recognition Technology and Privacy: Race and Gender - How to Ensure the Right to Privacy is Protected’ (2021) 23 San Diego International Law Journal 111; Joy Buolamwini and Timnit Gebru, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’ (2018) 81 Proceedings of Machine Learning Research 1.

[94] Lucía Vicente and Helena Matute, ‘Humans Inherit Artificial Intelligence Biases’ 13 Scientific Reports 1.

[95] Megan Garcia, ‘Racist in the Machine: The Disturbing Implications of Algorithmic Bias’ (2016) 33 World Policy Journal 111.

[96] Andrew D Selbst, ‘Disparate Impact in Big Data Policing’ (2017) 52 Georgia Law Review 109, 134-135.

[97] Nils Karl Reimer et al, ‘Self-Categorization and Social Identification: Making Sense of Us and Them’ in Derek Chadee (ed), *Theories in Social Psychology* (2nd edn, Wiley-Blackwell 2020) ; Miles Hewstone, Mark Rubin and Hazel Willis, ‘Intergroup Bias’ (2002) 53 Annual Review of Psychology 575, 576.

[98] Michael Bar-Eli et al, ‘Action Bias Among Elite Soccer Goalkeepers: The Case of Penalty Kicks’ (2007) 28 Journal of Economic Psychology 606.

سوگیری الگوریتمی در سامانه‌های تشخیص چهره (FRT) و دیگر سامانه‌های هوش مصنوعی قابل مشاهده است. برای نمونه، پروژه‌ی «سایه‌های جنسیت» در MIT نشان داد که الگوریتم‌های طبقه‌بندی جنسیت با داده‌هایی آموزش دیده بودند که عمدتاً شامل تصاویر افراد سفیدپوست بود؛ در نتیجه دقت‌شان در شناسایی افراد رنگین‌پوست کمتر بود. پژوهش‌ها نشان داده‌اند که خطای این الگوریتم‌ها برای زنان با پوست تیره تا ۳۴٪ بیشتر از مردان با پوست روشن است.[۹۹] این سوگیری می‌تواند احتمال شناسایی نادرست یا اشتباه گرفتن غیرنظامیان با اهداف قانونی را افزایش دهد.[۱۰۰] موارد واقعی هم نشان داده‌اند که به‌کارگیری سامانه‌های تشخیص چهره (FRT) در اجرای قانون، به‌طور نامتناسبی منجر به دستگیری افراد سیاه‌پوست شده است.[۱۰۱]

در عملیات هدف‌گیری، سوگیری الگوریتمی ممکن است توانایی سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) را در تمییز میان اهداف مشروع و غیرمشروع کاهش دهد.[۱۰۲] علاوه بر این، سوگیری انتساب گروهی مشکل‌ساز است: یک مطالعه در سال ۲۰۰۹ نشان داد که دانشجویان نظامی آمریکایی هنگام دیدن تصاویر «مردان خاورمیانه‌ای با پوشش سنتی» سریع‌تر شلیک کردند،[۱۰۳] زیرا آن‌ها را به‌طور کلی تروریست یا جنگجو فرض کرده بودند.[۱۰۴] در نهایت، سوگیری در عملکرد می‌تواند فشار روی فرماندهان برای تصمیم‌گیری سریع هنگام اتکا به توصیه‌های هوش مصنوعی را افزایش داده و احتمال خطا در هدف‌گیری را بالا ببرد.[۱۰۵]

۳.۴. ابهام (کدر بودن) هوش مصنوعی

سامانه‌های مبتنی‌بر هوش مصنوعی معمولاً برای توسعه‌دهندگان و کاربران «نیمه‌کدر» و ناشناخته هستند و فهم فرآیندها و نتایج آن‌ها دشوار است. انواع مختلفی از این ابهام وجود دارد که بر توانایی درک سیستم‌های هوش مصنوعی تأثیر می‌گذارد.[۱۰۶] شفافیت به فرآیندهای طراحی، آموزش، آزمایش و توسعه سامانه‌های هوش مصنوعی مربوط می‌شود، در حالی‌که قابلیت تفسیر به توانایی کاربران برای درک عملکرد سیستم و پیش‌بینی نتایج آن و همچنین توضیح چرایی تولید خروجی‌های خاص اشاره دارد.[۱۰۷] شکل دیگری از ابهام، قابلیت ردیابی است؛ یعنی توانایی کاربر یا توسعه‌دهنده برای دنبال کردن مسیر تصمیم‌گیری سیستم و بررسی دلایل توصیه‌های ارائه‌شده.[۱۰۸]

این انواع ابهام چالش‌هایی در تعامل انسان و ماشین ایجاد می‌کنند، امکان تأیید توصیه‌های تولیدشده توسط هوش مصنوعی را پیچیده می‌کنند و پرسش‌هایی درباره مسئولیت‌پذیری در جنایات جنگی مطرح می‌کنند. دولت‌ها و شرکت‌های خصوصی

---

[99] Buolamwini and Gebru (n 89).
[100] Kevin K Fleming, Carole L Bandy, and Matthew O Kimble, 'Decisions to Shoot in a Weapon Identification Task: The Influence of Cultural Stereotypes and Perceived Threat on False Positive Errors' (2010) 5 Social Neuroscience 201.
[101] Thaddeus L Johnson and Natasha N Johnson, 'Police Facial Recognition Technology Can't Tell Black People Apart' (*Scientific American*, 18 May 2023) <https://www.scientificamerican .com/article/police-facial-recognitiontechnology-cant-tell-black-people-apart/>.
[102] Ashley Deeks, 'Predicting Enemies' (2018) 104 Virginia Law Review 1529, 1577.
[103] Fleming, Bandy and Kimble (n 96).
[104] Keith B Payne and Joshua Correll, 'Race, Weapons, and the Perception of Threat', in Bertram Gawronski (ed), *Advances in Experimental Social Psychology* (Elsevier Academic Press 2020).
[105] Nema Milaninia, 'Biases in Machine Learning Models and Big Data Analytics: The International Criminal and Humanitarian Law Implications' (2020) 102 International Review of the Red Cross 199.
[106] It is important to note that these terms lack a universal definition, and their meanings can vary depending on the author.
[107] For more details about interpretability, see Jenna Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1, 1.
[108] Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 Columbia Law Review 1829, 1832; Arthur Holland Michel, 'The Black Box, Unlocked: Predictability and Understandability in Military ai' (United Nations Institute for Disarmament Research 2020).

معمولاً درباره داده‌های استفاده‌شده برای آموزش الگوریتم‌ها شفاف نیستند.[۱۰۹] نبود قابلیت تفسیر باعث می‌شود این سوال مطرح شود که آیا یک فرمانده نظامی می‌تواند توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) را درک کند، به ویژه اینکه چرا برخی اهداف به عنوان اهداف قانونی برچسب‌گذاری شده‌اند و دیگران نه.

درک توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) نیازمند دانش فنی بالا، مهارت تخصصی و منابع قابل‌توجه است.[۱۱۰] حتی برای توسعه‌دهندگان این سامانه‌ها. به ویژه برای کاربران، توصیه‌ها می‌توانند مانند «جعبه سیاه» عمل کنند.[۱۱۱] به دلیل فقدان قابلیت تفسیر، کاربران نمی‌دانند چرا افراد خاصی به عنوان اهداف پیشنهادی انتخاب شده‌اند. مشکلات مربوط به قابلیت ردیابی نیز می‌تواند در تحقیقات مربوط به جنایات جنگی چالش ایجاد کند.[۱۱۲] این وضعیت، به دلیل ابهام، پیچیدگی و غیرقابل پیش‌بینی بودن سامانه‌های مبتنی‌بر هوش مصنوعی، با عنوان « خلاء مسئولیت‌پذیری (Accountability Gap) » شناخته می‌شود.[۱۱۳]

سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) مبهم هستند زیرا توضیح نمی‌دهند چرا افراد خاصی به عنوان اعضای یک گروه مسلح توصیه شده‌اند. اگر توضیحاتی به کاربران داده نشود، خطر تصمیم‌گیری نادرست وجود دارد. برای یک فرمانده نظامی حیاتی است که بداند چه زمانی می‌تواند به توصیه‌های تولیدشده اعتماد کند، به ویژه وقتی پیامدهای عمل براساس آن توصیه‌ها ممکن است منجر به تلفات یا مرگ شود. همچنین، به دلیل فقدان قابلیت تفسیر، حتی توسعه‌دهندگان هوش مصنوعی ممکن است نتوانند پیش‌بینی کنند سیستم چگونه تصمیم‌گیری خواهد کرد، در چه رفتارهایی دخیل خواهد شد یا چه الگویی از داده‌ها استخراج خواهد شد.[۱۱۴] از آنجا که توصیه‌های الگوریتمی با سرعتی بسیار بالاتر از توانایی انسانی تولید می‌شوند، فرماندهان نظامی ممکن است تصمیمات فاجعه‌باری بگیرند که بالقوه با قوانین بین‌المللی بشردوستانه (IHL) در تضاد باشد.

۴. قانون انتخاب اهداف در عملیات نظامی

این بخش به کاربرد حقوق بین‌الملل بشردوستانه (IHL) و به ویژه قانون انتخاب اهداف در عملیات نظامی، در استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) و سامانه‌های تشخیص چهره (FRT) برای تصمیم‌گیری‌های هدفگیری نظامی می‌پردازد. قانون انتخاب اهداف در عملیات نظامی شامل قواعدی است که مشخص می‌کند تحت چه شرایطی افراد یا اهداف می‌توانند در درگیری‌های مسلحانه مورد حمله قرار گیرند. این بخش نشان می‌دهد چگونه سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) و سامانه‌های تشخیص چهره (FRT) می‌توانند با این قواعد همخوان یا ناسازگار باشند و از مطالعه موردی «لاوندر» برای تشریح این موضوع استفاده می‌کند. تمرکز این بخش بر سه اصل اصلی است:

---

[109] Brent Mittelstadt, 'Interpretability and Transparency in Artificial Intelligence' in Carissa Véliz (ed), *The Oxford Handbook of Digital Ethics* (Oxford University Press 2022).
[110] Burrell (n 103), 4.
[111] Yavar Bathaee, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 Harvard Journal of Law & Technology 890, 891.
[112] Marta Bo, Laura Bruun, and Vincent Boulanin, 'Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving aws' (sipri, October 2022).
[113] Marta Bo, 'Autonomous Weapons and the Responsibility Gap in Light of the *Mens Rea* of the War Crime of Attacking Civilians in the icc Statute' (2021) 19 Journal of International Criminal Justice 275.
[114] Bathaee (n 107), 924.

۴.۱. اصل تمایز

اصل تمایز توسط دیوان بین‌المللی دادگستری (ICJ) به عنوان «اصل اساسی» حقوق بین‌الملل بشردوستانه (IHL) شناخته شده[۱۱۵] و به عنوان یک قاعده عرفی بین‌المللی (CIL) مطرح است[۱۱۶]. این قاعده در ماده ۴۸ پروتکل الحاقی ۱ (AP 1) آمده است:

> «به منظور حفاظت از جمعیت غیرنظامی و اهداف غیرنظامی، طرفین درگیری باید در همه زمان‌ها بین جمعیت غیرنظامی و رزمندگان و بین اهداف غیرنظامی و اهداف نظامی تمایز قائل شوند و عملیات خود را فقط علیه اهداف نظامی هدایت کنند.»

این قاعده ایجاب می‌کند که طرفین درگیری مسلحانه بین اهداف قانونی و غیرقانونی تمایز قائل شوند و حملات خود را تنها به اهداف قانونی هدایت کنند. هدفگیری مستقیم غیرنظامیان یا جمعیت غیرنظامی ممنوع است[۱۱۷]، مگر آنکه یک غیرنظامی با مشارکت مستقیم در درگیری‌ها، حفاظت قانونی خود را از دست داده و به عنوان «غیرنظامی مشارکت‌کننده در نبرد» (DPH) شناخته شود[۱۱۸]. اهداف غیرنظامی نباید مورد حمله قرار گیرند و به صورت منفی تعریف می‌شوند؛ یعنی اشیایی که هدف نظامی محسوب نمی‌شوند[۱۱۹]. اهداف نظامی، آن دسته از اهدافی هستند که به دلیل ماهیت، موقعیت، هدف یا نحوه استفاده‌شان، سهم مؤثری در عمل نظامی دارند و تخریب، تصرف یا خنثی‌سازی آن‌ها در شرایط موجود، مزیت نظامی قابل‌توجهی فراهم می‌کند[۱۲۰]. این تعریف دو شرط انباشتی (cumulative) دارد که یکدیگر را تکمیل می‌کنند.

استفاده از سامانه‌های تشخیص چهره (FRT) می‌تواند شناسایی افراد را مؤثرتر و دقیق‌تر کند، البته بسته به سطح دقت داده‌های زیست‌سنجی موجود. اگر سامانه‌ای مانند «لاوندر» ورودی سامانه‌های تشخیص چهره (FRT) را دریافت کند، می‌تواند افراد ثبت‌شده در پایگاه داده زیست‌سنجی را جستجو، شناسایی و ردگیری کند. پژوهشگران معتقدند که استفاده از سامانه‌های تشخیص چهره (FRT) برای شناسایی افراد می‌تواند پایبندی به اصل تمایز را آسان‌تر کند[۱۲۱]. علاوه‌بر این، قوانین حقوق بین‌الملل بشردوستانه (IHL) به‌طور مستقیم استفاده از سامانه‌های تشخیص چهره (FRT) را ممنوع نمی‌کنند[۱۲۲]، اما ممکن است محدودیت‌هایی برای شناسایی گروه‌های خاصی از افراد وجود داشته باشد.[۱۲۳]

همان‌طور که پیش‌تر اشاره شد، ارتش اسرائیل در غزه به تشخیص چهره متکی بوده است. در زمینه درگیری‌های غزه، سامانه‌های تشخیص چهره (FRT) ممکن است برای شناسایی اینکه آیا یک فرد، مبارز شناخته‌شده حماس است یا نه، استفاده شود. سامانه‌های تشخیص چهره (FRT) تنها به شناسایی یا تأیید هویت فردی که قبلاً در پایگاه داده ثبت شده، کمک می‌کند و وضعیت قانونی او طبق حقوق بین‌الملل بشردوستانه (IHL) را تعیین نمی‌کند. برای مثال، ترکیب سامانه‌های پشتیبانی

---

115 *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) 1996 icj Rep. 226, 257.
116 Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Volume i: Rules* (cup 2005) 62, rule 7.
117 ap i (n 87), art 51; Henckaerts and Doswald-Beck (n 112), rule 1.
118 ap i (n 87), art 51(3) ; Henckaerts and Doswald-Beck (n 112), rule 6.
119 ap i (n 87), art 52(1) ; Henckaerts and Doswald-Beck (n 112), rule 8.
120 ap i (n 87), art 52.
121 Boothby (n 8), 397; Zwanenburg (n 43), 1416; Mitchell (n 7), 305-306.
122 Mitchell (n 7), 306.
123 Zwanenburg (n 43) ; Emily Crawford, 'The Right to Privacy and the Protection of Data for Prisoners of War in Armed Conflict' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (nato ccdcoe Publications 2022).

تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) با سامانه‌های تشخیص چهره (FRT) می‌تواند علاوه بر اطلاعات هویتی، نشان دهد که آیا فرد یک مبارز شناخته‌شده حماس است یا خیر.

با این‌حال، تضمین دقت سامانه‌های تشخیص چهره (FRT) در محیط‌های پیچیده و کنترل‌نشده دشوار است. در مناطق پرجمعیت مانند غزه، احتمال دارد سامانه به‌دلیل زاویه قرارگیری صورت نسبت به دوربین و سایه‌ها، چهره‌ها را به درستی شناسایی نکند. این موضوع منجربه مثبت‌های کاذب می‌شود، یعنی سامانه ممکن است به اشتباه یک غیرنظامی را به‌عنوان مبارز حماس شناسایی کند. به گفته یک مقام نظامی ارتش اسرائیل: «گاهی اوقات این فناوری غیرنظامیان را به اشتباه به عنوان اعضای حماس علامت‌گذاری می‌کرد»[۱۲۴]. بنابراین، استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) همراه با سامانه‌های تشخیص چهره (FRT) در مناطق پرجمعیت ممکن است کمتر مؤثر باشد و افزایش مثبت‌های کاذب می‌تواند به جمعیت غیرنظامی آسیب برساند.

اصل تمایز نیازمند تفسیر موقعیت‌محور (contextual) است تا مشخص شود چه کسی غیرنظامی است و چه کسی نیست و این به راحتی برای یک یک ماشین قابل ترجمه نیست. این ارزیابی نیازمند تحلیل پیچیده‌ای از اقدامات و رفتارهای افراد است تا مشخص شود که متعلق به غیرنظامیان یا مبارزان است.[۱۲۵] همان‌طور که کریستوف هینس (Christof Heyns) توضیح می‌دهد، سامانه باید بتواند بین «غیرنظامی که قطعه‌ای فلز در دست دارد» و «مبارزی با لباس غیرنظامی» تمایز قائل شود.[۱۲۶]

طبق حقوق بین‌الملل بشردوستانه (IHL)، هدفگیری غیرنظامیانی که به وضعیت DPH رسیده‌اند قانونی است، زیرا آن‌ها حفاظت قانونی خود را از دست داده‌اند[۱۲۷]. حقوق بین‌الملل بشردوستانه (IHL) دقیقاً تعریف نمی‌کند که DPH چه چیزی را شامل می‌شود، بلکه بیان می‌کند که غیرنظامیان از حملات مصون هستند «مگر زمانی‌که به‌طور مستقیم در درگیری‌ها مشارکت کنند»[۱۲۸]. این مورد به‌عنوان یک قاعده عرفی بین‌المللی پذیرفته شده است.[۱۲۹] با این‌حال، تعریف دقیق DPH همچنان مورد اختلاف پژوهشگران است[۱۳۰]. راهنمای تفسیری ICRC توضیحاتی برای تعیین زمان تبدیل یک غیرنظامی به DPH ارائه کرده است[۱۳۱]. ارزیابی DPH نیازمند تحلیل موقعیت‌محور سه عنصر سازنده است[۱۳۲] و این تحلیل یک ارزیابی کیفی است، نه عددی، که به قصد و شرایط فرد توجه دارد[۱۳۳]. درک این تحلیل و ترجمه آن به کدگذاری ماشینی برای دسته‌بندی،

---

[124] Frenkel (n 16).
[125] For more extensive details, see icrc, 'icrc Position on Autonomous Weapon Systems' (Geneva, 12 May 2021), 9.
[126] Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions' (Human Rights Council) A/hrc/23/47 (9 April 2013), 67.
[127] icrc, Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949 in Yves Sandoz, Christopher Swinarski and Bruno Zimmermann (eds) (1987) ('icrc Commentary'), para 1942.
[128] ap i (n 87), art 51(3).
[129] Henckaerts and Doswald-Beck (n 112), rule 6; The Supreme Court of Israel, *The Public Committee against Torture in Israel et al v. The Government of Israel et al* (Judgment) Case No. hcj 769/02 (11 December 2006), 50.
[130] Marco Sasso`li, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified' (2014) 90 International Law Studies 308; Michael N Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to Critics' (2013) Harvard National Security Journal Feature; Kenneth Watkin, 'Opportunity Lost: Organized Armed Groups and the icrc 'Direct Participation in Hostilities' Interpretive Guidance' (2010) 42 International Law and Politics 641.
[131] icrc, Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law (Geneva, 2009).
[132] For more details see ibid, 47-50, 51-58, 58-64.
[133] ibid, 42.

شناسایی الگوها و ترکیب تمام عناصر مرتبط به گونه‌ای که بتواند دقیق بین غیرنظامیان و DPH تمایز قائل شود، چالشی بزرگ است.[۱۳۴]

در موقعیت‌های هدفگیری، ارتش‌ها از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی استفاده می‌کنند تا داده‌ها را جمع‌آوری و اطلاعات را فیلتر کنند و بین اهداف قانونی و غیرقانونی تمایز قائل شوند. این تحلیل براساس تطابق الگو (pattern-matching) انجام می‌شود. به این ترتیب، برخی سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) بر فرض‌هایی متکی هستند که داده‌ها را به خروجی تبدیل می‌کنند[۱۳۵]. با این‌حال، این سامانه‌ها می‌توانند همبستگی‌ها را شناسایی کنند، اما قادر به تشخیص رابطه علی (causation) نیستند، زیرا این امر از نظر ریاضی قابل انجام نیست.[۱۳۶]

برای مثال، گفته شده است که سامانه «لاوندر» طراحی شده تا یاد بگیرد چگونه «ویژگی‌های اعضای شناخته‌شده حماس و PJJ را شناسایی کند».[۱۳۷] این مدل یادگیری ماشین با داده‌ها تغذیه شده تا احتمال اینکه یک فرد عضو حماس یا PJJ باشد را از ۱ تا ۱۰۰ رتبه‌بندی کند. یکی از افسران ارتش اسرائیل توضیح داده که لاوندر «گاهی افرادی را که الگوهای ارتباطی مشابه اعضای شناخته‌شده حماس یا PJJ داشتند، به اشتباه علامت‌گذاری کرده است».[۱۳۸] اگرچه به دلیل کمبود شفافیت در اسرائیل تأیید این اطلاعات دشوار است، اما آگاه بودن کاربران از این فرضیات ذاتی اهمیت حیاتی دارد تا از آسیب به غیرنظامیان جلوگیری شود و داده‌های مناسب برای تمایز بین اهداف قانونی و غیرقانونی استفاده شود.

برخی ارتش‌ها در چارچوب «حلقه OODA» (مشاهده، جهت‌گیری، تصمیم‌گیری، اقدام) عمل می‌کنند. با اتکا به سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS)، می‌توان این حلقه را سریع‌تر ایجاد و تصمیم‌گیری را تسریع کرد.[۱۳۹] گزارش‌ها حاکی از آن است که لاوندر برای تحلیل اطلاعات درباره اعضای بالقوه حماس یا دیگر گروه‌های مسلح در مرحله تعیین هدف استفاده می‌شود. اگرچه ارتش اسرائیل اعلام کرده که از لاوندر برای شناسایی یا پیش‌بینی تروریست بودن افراد استفاده نمی‌شود[۱۴۰]، اما گزارش‌ها نشان می‌دهند که این سامانه اطلاعات مختلف درباره افراد را جمع‌آوری و خروجی ارائه می‌دهد که نشان می‌دهد چه کسی ممکن است عضو حماس باشد.

موضوعی که به ویژه نگران‌کننده است، استفاده از سامانه «Where's Daddy?» است که مبارزان مشکوک را ردیابی می‌کند و وقتی وارد خانه خود می‌شوند، به ارتش اسرائیل هشدار می‌دهد. در برخی موارد، گزارش شده که ارتش اهداف را بدون تأخیر و حتی زمانی‌که اعضای خانواده هنوز در خانه بودند، بمباران کرده است.[۱۴۱] گفته می‌شود که هر دو سامانه، لاوندر و «Where's Daddy?»، در فرآیند هدفگیری استفاده شده‌اند. اگرچه غیرنظامیانی که وضعیت DPH دارند طبق حقوق بین‌الملل

---

134 Jeroen van den Boogaard, 'Proportionality and Autonomous Weapons Systems' (2015) 6 Journal of International Humanitarian Legal Studies 247, 262-263; Sasso`li (n 126), 328-330.
135 Holland Michel (n 4), 36-37.
136 Hengameh Irandoust and Abder Benaskeur, 'Human-Autonomy Teaming for Critical Command and Control Functions' (ieee International Conference on Human-Machine Systems, Rome, 2020), 3; Matteo Pasquinelli and Vladan Joler, 'The Nooscope Manifested: ai as Instrument of Knowledge Extractivism' (2021) 36 ai & Society 1263, 1276.
137 Abraham (n 12).
138 ibid.
139 Owen Daniels, 'Speeding Up the ooda Loop with ai: A Helpful or Limiting Framework?' 2021 Joint Air & Space Power Conference, 159.
140 idf Website (n 14).
141 Abraham (n 12).

بشردوستانه (IHL) اهداف قانونی محسوب می‌شوند، اما بمباران یک خانواده کامل بدون تأیید اینکه آیا هر عضو خانواده DPH است، به شدت نگران‌کننده است و احتمالاً با اصول تمایز، تناسب و احتیاط در حمله همخوانی ندارد.

همان‌طور که پیش‌تر گفته شد، خروجی لاوندر ابتدا توسط تحلیلگران اطلاعاتی بررسی می‌شود[۱۴۲]. در سال ۲۰۲۳، یک مقام ارتش اسرائیل توضیح داد که اکنون تحلیلگران انسانی نیازی ندارند یک هدف را ساعت‌ها بررسی کنند، زیرا بررسی آن تنها چند دقیقه طول می‌کشد[۱۴۳]. در صورت تأیید، اهداف بررسی‌شده به مسئولان برنامه‌ریزی و اجرای حملات منتقل می‌شوند. به گفته میمران (Mimran) و گال (Gal)، این اهداف به «اتاق هدف» فرستاده می‌شوند، جایی که مشاوران حقوقی، مشاوران عملیاتی و افسران ارشد اطلاعاتی آن‌ها را براساس اصول حقوق بین‌الملل بشردوستانه (IHL) بازبینی می‌کنند.[۱۴۴]

خروجی سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) وضعیت قانونی فرد را طبق حقوق بین‌الملل بشردوستانه (IHL) مشخص نمی‌کند، اما می‌تواند دسته‌بندی کاربر را راهنمایی کند. سامانه‌های تشخیص چهره (FRT) نیز تنها برای شناسایی یا تأیید هویت افراد براساس داده‌های زیست‌سنجی پیشین کاربرد دارند و نمی‌توانند وضعیت قانونی فرد را تعیین کنند. بنابراین، این مقاله استدلال می‌کند که کاربران وابسته به سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) و سامانه‌های تشخیص چهره (FRT) باید ارزیابی حقوقی انجام دهند تا وضعیت افراد طبق حقوق بین‌الملل بشردوستانه (IHL) مشخص شود. با توجه به احتمال نادیده گرفتن برخی منابع اطلاعاتی توسط سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS)، اهمیت دارد که پرسنل اطمینان حاصل کنند اهداف غیرقانونی از اهداف قانونی در مرحله انتخاب هدف قابل تمایز هستند. از آنجا که خروجی لاوندر به دقت و قابلیت اعتماد الگوریتم بستگی دارد، حیاتی است که پرسنل کافی برای بررسی دقت اهداف پیشنهادی و خروجی‌های سامانه‌های تشخیص چهره (FRT) وجود داشته باشند، در صورت نیاز اطلاعات اضافی دریافت شود و رعایت اصل تمایز تضمین گردد.

۴.۲. اصل تناسب

اصل تناسب (Proportionality) در حقوق بین‌الملل بشردوستانه (IHL) هم به عنوان یک ضمانت در برابر حمله به اهداف غیرقابل تمایز عمل می‌کند[۱۴۵] و هم یک اقدام احتیاطی است که افراد مسئول برنامه‌ریزی و تصمیم‌گیری درباره حملات را ملزم می‌سازد از اقداماتی که قاعده تناسب را نقض می‌کنند، خودداری کنند[۱۴۶]. این اصل یک قاعده عرفی بین‌المللی (CIL) محسوب می‌شود[۱۴۷] و در ماده ۵۱(۵) (ب) پروتکل الحاقی ۱ (AP I) به این شکل تدوین شده است:

> «حملاتی که احتمال تلفات یا جراحت غیرنظامیان، خسارت به اهداف غیرنظامی یا ترکیبی از آن‌ها را داشته باشد و این تلفات نسبت به مزیت نظامی مشخص و مستقیم موردانتظار بیش از حد باشد، ممنوع است.»

اصل تناسب ایجاب می‌کند که مسئولان عملیات قبل از حمله اثرات بالقوه آن، از جمله مرگ یا جراحت غیرنظامیان و تخریب اموال غیرنظامی را در نظر بگیرند. به این ترتیب، این اصل محدودیتی اضافی بر محدودیت ایجاد شده توسط اصل تمایز

---

142 ibid.
143 Newman (n 21).
144 Mimran and Dahan (n 22).
145 ap i (n 87), art 51(5) (b).
146 ap i (n 87), art 57(2) (a) (iii).
147 Henckaerts and Doswald-Beck (n 112), rule 14.

محسوب می‌شود. مسئولان موظف‌اند اثرات اتفاقی حملات برنامه‌ریزی‌شده -که به آن «خسارت جانبی» گفته می‌شود- را ارزیابی کنند و آن‌هایی که نسبت به مزیت نظامی بیش از حد هستند، شناسایی کنند. غیرنظامیان و اهداف غیرنظامی باید تا حد امکان از خسارت جانبی محفوظ بمانند. این اصل می‌پذیرد که خطر آسیب یا مرگ اتفاقی در جنگ وجود دارد و حمله نظامی می‌تواند قانونی باشد مادامی که خسارت جانبی موردانتظار بیش از مزیت نظامی ملموس و مستقیم موردانتظار نباشد.[۱۴۸]

مزیت نظامی موردنظر باید به صورت موقعیت‌محور (contextual) [۱۴۹] درک شود و واقعی و مشخص باشد، نه فرضی[۱۵۰]. دیدگاه‌های مختلفی درباره اینکه چه چیزی مزیت نظامی را تشکیل می‌دهد وجود دارد[۱۵۱]، اما تفسیر پذیرفته‌شده معمولاً «با در نظر گرفتن کل حمله» است، نه حملات جداگانه.[۱۵۲] مفهوم «بیش از حد» یا excessive نیز بسیار سخت و دشوار برای تعریف و اعمال است.[۱۵۳] عنصر کلیدی دیگر در ارزیابی تناسب، «آسیب اتفاقی» به غیرنظامیان است[۱۵۴] که شامل از دست دادن جان، آسیب روانی، تخریب اهداف غیرنظامی، آسیب به محیط زیست[۱۵۵] و اثرات موجی[۱۵۶] می‌شود. ماده ۵۷ (۲) (الف) (۳) پروتکل الحاقی ۱ نیز مشخص می‌کند که رعایت این اصل بر عهده کسانی است که حمله را برنامه‌ریزی یا تصمیم‌گیری می‌کنند. اگرچه «استاندارد فرمانده معقول» در حقوق بین‌الملل بشردوستانه (IHL) تعریف نشده، اما به عنوان معیاری برای قضاوت درباره تناسب توصیف شده است.[۱۵۷]

همان‌طور که پیش‌تر بحث شد، گزارش‌های مجله +۹۷۲ و Local Call نشان می‌دهند که سامانه لاوندر همراه با سامانه دیگری به نام the Gospel استفاده می‌شود تا تلفات موردانتظار در هر حمله محاسبه شود.[۱۵۸] تعداد تلفات پیش‌بینی‌شده- برآورد خسارت جانبی (CDE) - قبل از حمله تعیین می‌شود و واحدهای ارتش اسرائیل از آن مطلع می‌شوند. سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) که برآورد خسارت جانبی CDE ارائه می‌دهند، می‌توانند دقت تصمیم‌گیری را با متعادل کردن نیاز به مزیت نظامی و کاهش حداقل تلفات جانبی، افزایش دهند.[۱۵۹]

با این‌حال، یک نگرانی جدی این است که گزارش‌ها حاکی از آن است که ارتش اسرائیل «محدودیت‌ها در مورد تلفات غیرنظامیان موردانتظار را کاهش داده است».[۱۶۰] از یک سو، ارتش بر دقت بیشتر در هدفگیری تأکید می‌کند که با جمع‌آوری

---

148 Amichai Cohen and David Zlotogorski, *Proportionality in International Humanitarian Law: Consequences, Precautions, and Procedures* (Oxford University Press 2021), 4.
149 Yoram Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (3rd edn, Cambridge University Press 2016), 108.
150 Nils Melzer, *Targeted Killing in International Law* (Oxford University Press 2009), 293.
151 For a narrower interpretation see, icrc Commentary (n 123), para 2218. For broader interpretations see Judith Gardam, *Necessity, Proportionality and the Use of Force by States* (Cambridge University Press 2004), 101; Program on Humanitarian Policy and Conflict Research (hpcr), *Commentary to the hpcr Manual on International Law Applicable to Air and Missile Warfare* (Cambridge University Press 2013), 45.
152 Cohen and Zlotogorski (n 144), 66.
153 Michael Wells-Greco, 'Operation 'Cast Lead': *Jus in Bello* Proportionality' (2010) 57 Netherlands International Law Review 397, 399.
154 icrc Commentary (n 123), para 1913; Geoffrey Corn and Andrew Culliver, 'Wounded Combatants, Military Medical Personnel, and the Dilemma of Collateral Risk' (2017) 45 Georgia Journal of International and Comparative Law 445.
155 Cohen and Zlotogorski (n 144), 78-82.
156 For more details, see Ian Henderson and Kate Reece, 'Proportionality under International Humanitarian Law: The 'Reasonable Military Commander' Standard and Reverberating Effects' (2018) 51 Vanderbilt Journal of Transnational Law 835.
157 ibid, 840.
158 Abraham (n 12).
159 Sasso`li (n 126).
160 Abraham (n 23).

سریع و خودکار اطلاعات برای تولید اهداف ممکن شده است.[۱۶۱] ارتش اسرائیل بیانیه‌ای صادر کرده است مبنی‌بر اینکه پس از تأیید یک هدف برای حمله، آنها برای هر حمله ارزیابی فردی انجام می‌دهند.[۱۶۲] از ســـوی دیگر، یکی از مقامات ارتش گفته است که «در عمل، اصل تناسب رعایت نمی‌شود».[۱۶۳] مقام دیگری توضیح داده که در برخی حملات، ارتش صدها غیرنظامی را در پیگیری اهداف فرماندهان ارشد حماس مجاز دانسته است.[۱۶۴]

به‌طور کلی، استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) می‌تواند پایبندی به اصل تناسب را با ارائه اطلاعات دقیق‌تر درباره تعداد بالقوه تلفات بهبود بخشـــد. با این‌حال، نگرانی‌های قابل‌توجهی وجود دارد مبنی‌بر اینکه آیا ارتش به‌طور مؤثر از اطلاعات موجود برای ارزیابی خسارت جانبی استفاده می‌کند و تضمین می‌کند که آن نسبت به مزیت نظامی موردانتظار متناسب باقی بماند یا خیر.

با این‌حال پژوهشگران در تعریف مفهوم تناسب و اینکه سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) تا چه حد می‌تواند در ارزیابی‌ها مفید باشـــد اختلاف نظر دارند. برخی معتقدند مشـــخص نیســـت که چگونه الزامات پیچیده، موقعیت‌محور و مبتنی‌بر ارزش را می‌توان در قالب یک فرمول ریاضی پیاده‌سازی کرد.[۱۶۵] دیگران می‌گویند با توجه به پیشرفت سریع فناوری، ممکن است بتوان اجرای اصل تناسب را در کدگذاری ماشین پیش‌برنامه‌ریزی کرد.[۱۶۶] همچنین بحث شده که حقوق بین‌الملل بشـردوسـتانه (IHL) نیازمند قضـاوت‌های ذهنی نیسـت که ماشـین‌ها قادر به انجام آن نباشـند، بلکه بر ارزیابی عینی واقعیت‌ها متکی است.[۱۶۷] با این‌حال، برخی دیگر این سوال را مطرح می‌کنند که آیا الگوریتم‌های مورداسـتفاده در جنگ قادر به متعادل کردن آسـیب‌های تصـادفی در برابر مزیت نظامی پیش‌بینی‌شـده هسـتند یا خیر.[۱۶۸] علاوه‌بر این، مزیت نظامی و آسـیب‌های تصـادفی به غیرنظامیان «را نمی‌توان از طریق اسـتفاده سـاده از یک فرمول مقایسـه کرد، زیرا هیچ مخرج مشـترکی بین آنها وجود ندارد».[۱۶۹] به‌طور خلاصه، دشواری اصلی در «کمی‌سازی عوامل معادله»[۱۷۰] است، زیرا این فرآیند یک استاندارد حقوقی ذهنی و نامعین دارد.[۱۷۱] همان‌طور که یورام دینشتاین می‌گوید:

> « مزیت نظامی و تلفات یا خسـارت وارده بر غیرنظامیان از نظر کمّی با هم قابل مقایسـه نیسـتند و نمی‌توان آن‌ها را طوری تنظیم کرد که در یک مخرج عددی مشـترک قرار گیرند. تلفات پیش‌بینی‌شـده و خسـارات وارده بر غیرنظامیان را می‌توان

---

[161] ۱۵۷ הצצה [A Glimpse of the idf’s Target Factory that Operates Around the Clock] (*Israel Defense Forces*, 2 November 2023) <https://www.idf.il/%D7%90%D7%AA%D7%A8%D7%99%D7%99%D7%97%D7%99%D7%93%D7%95%D7%AA/%7%99%D7%95%D7%9E%D7%9F-%D7%94%D7%9E%D7%9C%D7%97%D7%9E%D7%94/%D7%9B%D7%9C-%D7%94%D7%9B%D7%AA%D7%91%D7%95%D7%AA/%D7%94%D7%A4%D7%A6%D7%95%D7%AA/%D7%9E%D7%9C%D7%97%D7%9E%D7%94-%D7%9E%D7%98%D7%A8%D7%95%D7%AA-%D7%A9%D7%94%D7%95%D7%AA%D7%A7%D7%A4%D7%95-%D7%9B%D7%95%D7%97%D7%95%D7%AA-%D7%A6%D7%94-%D7%9C-%D7%90%D7%92%D7%A3%D7%94%D7%9E%D7%95%D7%93%D7%99%D7%A2%D7%99%D7%9F-%D7%97%D7%99%D7%9C-%D7%94%D7%90%D7%95%D7%95%D7%99%D7%A8-%D7%97%D7%99%D7%9C-%D7%94%D7%99%D7%9D/>.

[162] idf Website (n 14).

[163] Abraham (n 12).

[164] bid.

[165] Robert D Sloane, ‘Puzzles of Proportion and the Reasonable Military Commander: Reflections on the Law, Ethics, and Geopolitics of Proportionality’ (2015) 6 Harvard National Security Journal 299, 322-323; Cohen and Zlotogorski (n 144), 59.

[166] Schmitt (n 126) ; Sasso`li (n 126).

[167] Sasso`li (n 126), 339.

[168] van den Boogaard (n 130), 267.

[169] Cohen and Zlotogorski (n 144), 59.

[170] Michael Bothe, Karl Joseph Partsch and Waldemar Solf, *New Rules for Victims of Armed Conflicts Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (2nd edn, Martinus Nijhoff Publishers 2013), 227.

[171] Sloane (n 161), 301-302.

محاسبه یا برآورد کرد؛ اما چگونه می‌توان مزیت نظامی مورد انتظار را به‌صورت معیاری قابل اندازه‌گیری سنجید؟ همین غیرقابل‌مقایسه‌بودن اغلب هرگونه تلاش برای انجام یک سنجش عینی و متوازن بین این دو را بی‌اثر می‌کند..»[۱۷۲]

من استدلال می‌کنم که ارزیابی‌های تناسب ماهیت ذهنی دارند و تحت تأثیر طیف گسترده‌ای از قضاوت‌ها قرار می‌گیرند.[۱۷۳] این ارزیابی براساس سنجش دو ارزش متفاوت انجام می‌شود: (۱) مزیت نظامی موردانتظار و (۲) تلفات یا خسارت غیرنظامی پیش‌بینی‌شده. در حالی‌که تعداد تلفات غیرنظامیان قابل برآورد است، این تنها بخشی از آزمون تناسب است. مطابق پروتکل الحاقی ۱، فرد مسئول باید ارزیابی کند که آیا خسارت یا جراحت غیرنظامیان «نسبت به» مزیت نظامی مشخص و مستقیم، بیش از حد (excessive) خواهد بود یا خیر. بنابراین، تعیین یک عدد از پیش‌تعیین‌شده برای تلفات غیرنظامی نمی‌تواند به‌طور دقیق اصل تناسب را ارزیابی کند. یک حمله نمی‌تواند این اصل را نقض کند، مگر اینکه خسارات جانبی در مقایسه با مزیت نظامی ملموس و مستقیم، بیش از حد باشد. ایجاد تعادل بین این دو مفهوم، یک قضاوت ذهنی است، نه عینی.[۱۷۴] به دلیل پیچیدگی این مفاهیم و به دلیل اینکه تعریف روشنی از آنها در بین محققان وجود ندارد، مشخص نیست که آیا و چگونه می‌توان این مفاهیم را در کدگذاری ماشینی تبدیل و عملیاتی کرد. علاوه بر این، تفسیر کمیته بین‌المللی صلیب سرخ نشان می‌دهد که اگرچه «تا حدودی مبتنی‌بر ارزیابی ذهنی است، تفسیر [اصل تناسب] بیش از هر چیز باید مسئله‌ای مبتنی‌بر عقل سلیم و حسن نیت برای فرماندهان نظامی باشد.»[۱۷۵]

در نهایت، بار اصلی ارزیابی اثرات حملات پیشنهادی برعهده فرمانده نظامی است، نه سامانه هوش مصنوعی و او باید به دقت منافع بشردوستانه و نظامی را در نظر بگیرد[۱۷۶]، زیرا این فرمانده است که شخصاً مسئول پایبندی به اصل تناسب است. بنابراین، فرمانده نظامی باید از امکان خطا یا شکست سامانه‌های هوش مصنوعی آگاه باشد و نباید تصور کند که خروجی‌های آن‌ها حقیقت مطلق یا داده‌های بی‌خطا را ارائه می‌دهند.[۱۷۷] برای مثال، اگر سامانه «the Gospel» خسارت جانبی را کمتر از واقعیت برآورد کند، این مسئولیت فرمانده نظامی است که صرف‌نظر از صحت یا عدم صحت اطلاعات ارائه شده توسط سامانه the Gospel، ظرفیت انجام ارزیابی‌های صحیح را حفظ کند.

این مسئله پرسشی ایجاد می‌کند: چگونه می‌توان استاندارد «فرمانده نظامی معقول» را در تعامل با توصیه‌های تولیدشده توسط الگوریتم‌ها اعمال کرد. در گزارش نهایی کمیته دادگاه بین‌المللی کیفری یوگسلاوی سابق در سال ۲۰۰۰ آمده است که «تعیین ارزش‌های نسبی باید بر عهده فرمانده نظامی معقول باشد».[۱۷۸] همچنین، در پرونده گالیچ، دادگاه بین‌المللی کیفری برای یوگسلاوی سابق اعلام کرد:

---

[172] Dinstein (n 145), 158.
[173] Yuval Shany, 'Toward a General Margin of Appreciation Doctrine in International Law?' (2005) 16 European Journal of International Law 907; Luke Whittemore, 'Proportionality Decision Making in Targeting: Heuristics, Cognitive Biases, and the Law' (2016) 7 Harvard National Security Journal 577.
[174] Cohen and Zlotogorski (n 144), 59.
[175] icrc Commentary (n 123), para 2208.
[176] ibid.
[177] US DoD (n 66).
[178] Office of the Prosecutor, 'International Criminal Tribunal for the Former Yugoslavia: Final Report to the Prosecutor by the Committee Established to Review the nato Bombing Campaign against the Federal Republic of Yugoslavia' ('Final Report') (2000) 39 International Legal Materials 1257, 50.

«برای تعیین اینکه آیا یک حمله متناسب بوده است یا خیر، لازم است بررسی شود که آیا یک شخص نسبتاً آگاه در شـرایط واقعی مشـابه و با اسـتفاده معقول از اطلاعات موجود می‌توانسـته پیش‌بینی کند که تلفات غیرنظامیان بیش از حد خواهد بود یا خیر.»[۱۷۹]

همان‌طور که گفته شـد، فرمانده نظامی نباید «چشـم بر حقایق ببندد»، بلکه موظف اسـت تمام اطلاعات موجود را در نظر بگیرد.[۱۸۰] علاوه بر این، آنچه ممکن اسـت برای یک فرمانده معقول باشـد، ممکن اسـت برای دیگری غیرمعقول تلقی شـود. اختلاف نظر در این ارزیابی‌ها ممکن اسـت مسـائل مربوط به انتسـاب مسـئولیت کیفری به فرماندهان نظامی را پیچیده کند. به‌عنوان مثال، مطالعه‌ای در سال ۲۰۲۰ نشان داد که کارشناسان آکادمیک و نظامی در مورد تعیین حداکثر تلفات قابل قبول غیرنظامیان توافق نداشـتند.[۱۸۱] تحقیقات قبلی نیز نشـان داده‌اند که عوامل انسـانی فردی، مانند پس‌زمینه و ارزش‌های تصمیم‌گیرنده[۱۸۲]، تجربه رزم و سابقه نظامی ملی[۱۸۳]، می‌توانند بر قضاوت «فرمانده نظامی معقول» اثرگذار باشند.

من معتقدم که عوامل خارجی متعددی وجود دارند که می‌توانند تصمیم فرمانده نظامی را شکل دهند و تحت تأثیر قرار دهند. اولین عامل، طراحی و استفاده از توصیه‌های الگوریتمی است که اهداف قانونی را شناسایی و پیشنهاد می‌کنند. این توصیه‌ها اطلاعاتی ارائه می‌دهند که آگاهی موقعیتی فرماندهان را درباره میدان نبرد شـکل می‌دهند و در عرض چند ثانیه در زمان واقعی تولید می‌شـوند. حجم زیاد توصـیه‌های روزانه فشـار بر فرماندهان برای عمل را افزایش می‌دهد و امکان تصـمیمات شتاب‌زده را بالا می‌برد، زیرا زمان محدودی برای انجام ارزیابی تناسب باقی می‌ماند.

عامل دوم، ابهام هوش مصـنوعی اسـت. اگر تحلیلگران یا فرماندهان نتوانند بفهمند چرا سـامانه‌های پشـتیبانی تصـمیم‌گیری مبتنی‌بر هوش مصـنوعی (AI-DSS) یک فرد را به‌عنوان هدف قانونی طبقه‌بندی کرده اسـت و ماهیت هدف را تأیید کنند، تصـمیم‌گیری کاملاً مطلعانه دشـوار می‌شـود. سـوم، ممکن اسـت دانش و مهارت فنی محدود برای تفسـیر و اسـتفاده از هوش مصـنوعی برای اطلاع‌رسـانی به تصـمیمات فرماندهان وجود داشـته باشـد. سـؤال کلیدی این اسـت که آیا فرمانده واقعاً «به‌اندازه کافی مطلع و توانمند» اسـت تا در اسـتفاده معقول از اطلاعات موجود، اتکا به توصـیه‌های الگوریتمی، ابهام سـامانه و محدودیت‌های فنی را در نظر بگیرد یا خیر.

من معتقدم که پاسخ به این پرسش بستگی دارد به این که آیا یک فرمانده نظامی قادر است از اطلاعات موجود در توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) به‌صورت معقول استفاده کند یا خیر. سطح درک فنی موردنیاز برای فرماندهان ممکن است بسته به نقش، سطح عملیاتی و کارشناسان اطلاعاتی پشتیبان آن‌ها متفاوت باشد. با این‌حال، حقوق بین‌الملل بشردوستانه (IHL) سطح دانش فنی فرماندهان نظامی را مشخص نمی‌کند؛ یعنی تعیین نمی‌کند چه آموزش فنی باید ببینند، چه دانش فنی بعدی باید کسب کنند و چگونه از نرخ خطا و دلایل احتمالی شکست مطلع شوند.[۱۸۴] با توجه به پیشـرفت‌های هوش مصـنوعی نظامی، احتمالاً سـامانه‌ها پیچیده‌تر و هوشـمندتر خواهند شـد و ابهام آن‌ها افزایش

---

[179] *Prosecutor v Galic* (Trial Chamber) it-98-29 (5 December 2003), [58].
[180] Frits Kalshoven, 'Implementing Limitations on the Use of Force: The Doctrine of Proportionality and Necessity', (1992) 86 Proceedings of the Annual Meeting (American Society of International Law) 39, 44.
[181] Daniel Statman et al, 'Unreliable Protection: An Experimental Study of Experts' *In Bello* Proportionality Decisions' (2020) 31 European Journal of International Law 429.
[182] Final Report (n 174), 50.
[183] ibid.
[184] For further discussion, see Jonathan Kwik, 'Lawfully Using Autonomous Weapon Technologies: A Theoretical and Operational Perspective' (PhD Thesis, University of Amsterdam [2024]), ch 5.

می‌یابد.[۱۸۵] گفته شده است که «ممکن است برخی فناوری‌ها هرگز به سطح شفافیتی که توسط قانون‌گذاران و دولت‌ها مطلوب است، نرسند».[۱۸۶] بنابراین، لازم است بررسی شود که استاندارد «فرمانده نظامی معقول» در این زمینه چه معنایی دارد و تا چه حد فرماندهان می‌توانند در تصمیم‌گیری‌های هدفگیری به توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) تکیه کنند.

۴.۳. اصل احتیاط در حمله

اصل احتیاط در حمله در ماده ۵۷ پروتکل الحاقی ۱ تدوین شده و به‌عنوان یک قاعده عرفی بین‌المللی پذیرفته شده است.[۱۸۷] این اصل ایجاب می‌کند که در اجرای عملیات نظامی باید اقدامات احتیاطی برای حفاظت از غیرنظامیان و اهداف غیرنظامی اتخاذ شود. اصطلاح «عملیات نظامی» شامل هرگونه حرکت، مانور و فعالیت‌های مرتبط با نبرد است که توسط نیروهای مسلح انجام می‌شود.[۱۸۸] ماده ۵۷(۱) همچنین تأکید می‌کند که طرفین درگیری باید در همه زمان‌ها «مراقبت مستمر» را برای محافظت از غیرنظامیان به عمل آورند؛ این تعهد، تعهدی مستمر بوده و هیچ محدودیت زمانی ندارد.[۱۸۹]

استدلال شده است که وظیفه مراقبت مستمر یک «تعهد کلی، گسترده و انعطاف‌پذیر» است[۱۹۰] و محدود به موارد خاص ذکرشده در ماده ۵۷ نیست.[۱۹۱] بنابراین، برخی پژوهشگران معتقدند این ماده به‌خودی‌خود یک تعهد حقوقی ایجاد می‌کند[۱۹۲] و دامنه کاربرد آن گسترده‌تر است.[۱۹۳] در پاراگراف اول ماده ۵۷، این تعهد به «عملیات نظامی» اعمال می‌شود که شامل هر حرکت یا فعالیتی است که با نیت نبرد انجام می‌شود،[۱۹۴] در حالی‌که پاراگراف‌های بعدی به «حمله» اشاره دارند که دامنه محدودتری دارد. به نظر می‌رسد که «مراقبت مستمر» دامنه حفاظتی گسترده‌تری دارد، زیرا شامل فعالیت‌هایی فراتر از حملات مستقیم نیز می‌شود. اصطلاح «مراقبت مستمر» توسط حقوق بین‌الملل بشردوستانه (IHL) تعریف نشده است، اما کتاب راهنمای تالین ۲.۰ توضیح می‌دهد که این تعهد به این معناست که فرماندهان و همه افراد درگیر باید به‌طور مداوم نسبت به اثرات فعالیت‌های خود بر غیرنظامیان و اهداف غیرنظامی حساس باشند و تلاش کنند از هرگونه اثر غیرضروری جلوگیری کنند.[۱۹۵]

یک تفسیر مشابه می‌تواند برای سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) نیز مطرح شود. سوگیری‌های فنی ذاتی در سامانه‌های هوش مصنوعی از مرحله توسعه می‌تواند باعث آسیب فیزیکی شود. برای مثال، اگر سامانه هدفگیری مبتنی‌بر هوش مصنوعی عمدتاً با چهره‌های سفیدپوست آموزش دیده باشد، احتمال خطای آن برای افراد رنگین‌پوست بیشتر است و این می‌تواند منجر به آسیب فیزیکی به آن‌ها شود. پژوهش‌های پیشین نشان داده‌اند که نرم‌افزار

---

185 Bathaee (n 107), 929.
186 ibid.
187 Henckaerts and Doswald-Beck (n 112), rule 15.
188 icrc Commentary (n 123), para 2191.
189 Asaf Lubin, 'The Duty of Constant Care and Data Protection in War' in Laura A Dickinson and Edward Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023).
190 ibid, 236.
191 Eric Talbot Jensen, 'Cyber Attacks: Proportionality and Precautions in Attack' (2013) 89 International Law Studies 198, 202.
192 ibid; Jean-François Quéguiner, 'Precautions under the Law Governing the Conduct of Hostilities' (2006) 88 International Review of the Red Cross 793.
193 Quéguiner (n 189).
194 icrc Commentary (n 123), para 2191.
195 Michael N Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge University Press 2017), 477.

تشخیص چهره در شناسایی افراد سفیدپوست بهتر عمل کرده و نرخ شناسایی صحیح تا ۹۹٪ برای مردان سفیدپوست بوده است، در حالی که برای زنان با پوست تیره به ۳۴٪ کاهش یافته است، زیرا الگوریتم‌ها عمدتاً با چهره‌های مردانه سفیدپوست آموزش دیده بودند.[۱۹۶]

بنابراین، ضروری است که سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) که ورودی خود را از سامانه‌های تشخیص چهره (FRT) دریافت می‌کنند و در درگیری‌های مسلحانه استفاده می‌شوند، براساس داده‌هایی باشند که نماینده جمعیت هدف واقعی هستند. در حال حاضر، شفافیت کافی در مورد توسعه هوش مصنوعی نظامی وجود ندارد و معلوم نیست سوگیری نژادی در الگوریتم‌ها تا چه حد وجود دارد.[۱۹۷]وظیفه «مراقبت مستمر» ممکن است ایجاب کند که کشورها تلاش کنند تا مطمئن شوند داده‌های زیست‌سنجی و اهداف از پیش انتخاب‌شده دقیق و به‌روز هستند و این بررسی باید قبل، حین و به‌طور مستمر در طول عملیات نظامی انجام شود.

علاوه بر این، این تعهد ممکن است ایجاب کند که توسعه‌دهندگان هوش مصنوعی نظامی تلاش کنند تا داده‌های آموزشی نماینده و جامع جمع‌آوری کنند، به‌گونه‌ای که سوگیری و خروجی‌های غیرقابل پیش‌بینی به حداقل برسد و جمعیت غیرنظامی در برابر آسیب، جراحت و از دست دادن جان محافظت شود. الزام خاص برای تأیید اهداف، همانطور که در ماده ۵۷(۲) (الف) (۱) پروتکل الحاقی ۱ آمده است، ممکن است نیازمند نظارت مستمر بر داده‌ها برای اطمینان از دقت باشد. برای تضمین عملکرد صحیح سامانه‌های هدفگیری، ضروری است که این سامانه‌ها با داده‌های نماینده آموزش داده شوند.

با این‌حال، ماهیت الزام‌آور وظیفه مراقبت مستمر هنوز نیاز به روشن‌سازی دارد، زیرا مشخص نیست این وظیفه چگونه بر استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) در درگیری‌های مسلحانه اعمال می‌شود. طبق گزارش مجله +۹۷۲ و Local Call، الگوریتم‌های یادگیری ماشین با داده‌های آموزشی‌ای تغذیه شدند که شامل اطلاعاتی درباره کارکنان غیرنظامی دولت حماس بود، از جمله پلیس، اعضای دفاع شهری، بستگان militants و حتی ساکنانی از غزه که نام مشابه با یک عضو شناخته‌شده حماس داشتند. این باعث شد که لاوندر به‌طور نادرست برخی غیرنظامیان را به عنوان عوامل حماس علامت‌گذاری کند، زیرا الگوهای ارتباطی یا رفتاری مشابه با عوامل شناخته‌شده داشتند.[۱۹۸]

علاوه بر این، در سال ۲۰۲۱ گزارش شد که یکی از چالش‌های IDF در استقرار سامانه این بود که داده کافی برای آموزش الگوریتم‌ها بر روی آنچه که هدف نیست، وجود نداشت.[۱۹۹] این موضوع نگران‌کننده است، زیرا نشان می‌دهد الگوریتم‌ها با داده‌های نماینده توسعه نیافته‌اند و تنها نمونه‌هایی از اهداف مشروع (lawful) برای آموزش استفاده شده‌اند. پیامد این نمونه‌گیری، ایجاد خطاهای سیستماتیک و توصیه‌های نادرست است، زیرا سامانه نمی‌تواند به درستی بین افراد محافظت‌شده و غیرمحافظت‌شده تمایز قائل شود.[۲۰۰]

همانطور که اشاره شد، تمام مدل‌های تصادفی دارای یک حاشیه خطای مشخص هستند. گزارش شده است که لاوندر حدود ۱۰٪ حاشیه خطا دارد. بنابراین، از بین ۳۷۰۰۰ فردی که به عنوان اهداف نظامی علامت‌گذاری شده بودند، حدود ۳۷۰۰ نفر

---

196 Buolamwini and Gebru (n 89).
197 ibid.
198 Abraham (n 12).
199 Gaza Conflict Task Force, 'Gaza Conflict 2021 Assessment: Observations and Lessons' (October 2021) <https://jinsa.org/wp-content/uploads/2021/10/Gaza-Assessment.v8-1.pdf> accessed 1 February 2024, 31.
200 Selbst (n 92), 134-135.

در واقع غیرنظامی بوده‌اند.[۲۰۱] در طول درگیری، IDF تعریف «یک عامل حماس» را تغییر داد و دامنه آن را گسترش داد.[۲۰۲] این باعث نگرانی‌هایی درباره داده‌های استفاده‌شده برای برچسب‌زنی اهداف شده است. IDF اعلام کرده که مجموعه داده‌ها به‌طور منظم به‌روزرسانی و صحت‌سنجی می‌شوند[۲۰۳] و توسعه‌دهندگان در طول درگیری داده‌ها را مجدداً برچسب‌زنی و بهبود می‌کنند تا عملکرد الگوریتم ارتقا یابد.

با این‌حال، مگر اینکه کاربران کنونی سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) درباره چگونگی آموزش الگوریتم‌ها و برچسب‌های استفاده‌شده آگاه باشند، کدری سامانه برای فرماندهان نظامی تشدید می‌شود و این می‌تواند بر توانایی آن‌ها در درک دلیل توصیه الگوریتم برای هدف قرار دادن افراد خاص و تصمیم‌گیری براساس آن تأثیر بگذارد. این مثال خطر ذاتی برای جمعیت غیرنظامی از وجود حاشیه خطای بالا در سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) و نقش توسعه‌دهندگانی که این الگوریتم‌ها را طراحی، آموزش و به‌روزرسانی می‌کنند- که اغلب برای کاربران نهایی ناشناخته، نامرئی و اطلاع‌رسانی نشده است- نشان می‌دهد.

این مثال اهمیت حداقل کردن تأثیر کدری هوش مصنوعی و به‌روزرسانی دقیق برچسب‌ها در طول درگیری‌های مسلحانه را روشن می‌کند. برای درک کامل پیامدهای وظیفه مراقبت مستمر، ضروری است دامنه و کاربرد آن برای استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) به دقت بررسی شود.

ماده ۵۷(۲) (الف) (۱) پروتکل الحاقی ۱ ایجاب می‌کند که برنامه‌ریزان و تصمیم‌گیرندگان تمام اقدامات ممکن را انجام دهند تا مطمئن شوند که هدف یک هدف نظامی است و از حمایت ویژه برخوردار نیست. سامانه لاوندر با سرعت بی‌سابقه‌ای عمل می‌کند و توانایی دارد حجم عظیمی از اطلاعات را به سرعت پردازش کند. اثرات سوگیری اتوماسیون در چنین شرایطی ممکن است این باشد که لغو یا اصلاح حملات، با توجه به سرعت بالای تولید توصیه‌ها توسط فناوری، برای کاربران دشوارتر شود.

علاوه بر این، گفته شده است که سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) می‌تواند حجم زیادی داده را بهتر و سریع‌تر از انسان پردازش کرده و آن را به اهدافی برای حمله تبدیل کند.[۲۰۴] در نتیجه، احتمال دارد که تحلیلگران انسانی و فرماندهان نظامی به این محاسبات بیش از قضاوت خودشان اعتماد کنند و در جستجوی اطلاعات متناقض کوتاهی کنند. به همین دلیل، من تأکید می‌کنم که کاربران سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) باید زمان کافی برای بررسی توصیه‌های الگوریتمی صرف کنند و فرآیند تصمیم‌گیری را کند کنند تا اطمینان حاصل شود که اهداف غیرنظامی در برابر حمله مستقیم محافظت می‌شوند. با این‌حال، هنوز روشن نیست که چه میزان زمان باید به بررسی دقت توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) اختصاص داده شود تا رعایت اصل احتیاط تضمین شود.

نگرانی دیگر این است که چقدر زمان اختصاص‌یافته برای بررسی اهداف می‌تواند بر رعایت اصل احتیاط در حمله تأثیر بگذارد. در مراحل اولیه جنگ در اکتبر ۲۰۲۳، یک افسر اطلاعاتی توضیح داد که برای هر هدف حدود ۲۰ ثانیه «قبل از صدور مجوز بمباران» اختصاص می‌دادند[۲۰۵] و روزانه ده‌ها بررسی انجام می‌شد.[۲۰۶] گزارش‌ها نشان می‌دهد که بررسی‌ها گاهی تنها محدود

---

201 Abraham (n 12).
202 ibid.
203 idf Website (n 14).
204 Abraham (n 23).
205 ibid.
206 Abraham (n 12).

به تعیین جنسیت فرد بوده است. IDF اطلاعاتی در مورد مدت زمان اختصاص‌یافته برای بررسی حقوقی اهداف قبل از درگیری ارائه نکرده است.

اگر فرآیند بررسی توصیف‌شده در عمل اجرا می‌شد، حتی به صورت گاه‌به‌گاه، این نگرانی‌های جدی در زمینه حقوق بین‌الملل بشردوستانه (IHL) ایجاد می‌کند. نخست اینکه بررسی‌ها شامل ارزیابی کامل مشروعیت هدف تحت حقوق بین‌الملل بشردوستانه (IHL) نبوده، بلکه تنها تحلیلی از جنسیت فرد انجام شده است. دوم، نحوه و مسئولیت تحلیلگران اطلاعات در مرحله بازبینی اهداف نامشخص است.[۲۰۷] سوم، خطر عدم صرف زمان کافی برای تأیید اهداف وجود دارد.

اگر کاربران به توصیه‌های سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) متکی هستند، باید ارزیابی دقیقی از صحت و دقت هر توصیه انجام شود. آن‌ها باید با منابع و داده‌هایی که برای تولید این اهداف استفاده شده‌اند آگاه باشند تا در صورت نیاز، اطلاعات مرتبط دیگری جمع‌آوری کنند. در نهایت، باید یک تحلیل حقوقی انجام شود تا مشخص شود آیا فرد موردنظر، برای مثال وضعیت DPH دارد و می‌تولند به عنوان هدف مشروع (lawful) در نظر گرفته شود یا خیر. مسئولیت اولیه اجرای اقدامات احتیاطی بر عهده فرماندهان نظامی است، زیرا سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) جایگزین تصمیم‌گیرنده نمی‌شوند، بلکه صرفاً به تصمیم‌گیری کمک می‌کنند.[۲۰۸]

استفاده از فناوری تشخیص چهره (FRT) می‌تواند در تأیید هویت افراد کارآمد باشد و حتی می‌توان استدلال کرد که کسانی که حملات را برنامه‌ریزی و اجرا می‌کنند، موظف‌اند از سامانه‌های تشخیص چهره (FRT) استفاده کنند، اگر این فناوری در دسترسشان باشد تا «تمام اقدامات ممکن» را برای تأیید هویت اهداف خود انجام دهند.[۲۰۹] بنابراین، استفاده از سامانه‌های تشخیص چهره (FRT) در تصمیم‌گیری‌های هدفگیری ممکن است بحث‌برانگیز نباشد، بلکه در برخی موارد ضروری باشد.[۲۱۰] با این‌حال، سامانه‌های تشخیص چهره (FRT) تنها هویت فرد را شناسایی می‌کند و نمی‌تواند تعیین کند که یک شخص غیرنظامی است یا عضو یک گروه مسلح؛ بنابراین، کاربرد آن بیشتر برای شناسایی افراد مشخص شناخته‌شده مفید است تا در عملیات رزمی بزرگ و مناطق پرجمعیت.

همچنین، لازم است که هنگام بررسی توصیه‌های تولیدشده توسط سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) زمان کافی صرف شود، زیرا سامانه‌هایی مانند لاوندر اهداف را بسیار سریع‌تر از توان انسانی شناسایی می‌کنند. افزایش تعداد اهداف ممکن است سوگیری عملی را تشدید کند، به‌ویژه اگر استراتژی IDF بر کمیت بیش از کیفیت تأکید داشته باشد[۲۱۱] و محدودیت‌های خسارت جانبی کاهش یافته باشد که نشان‌دهنده آن است که اقدام به جای عدم اقدام با استراتژی کلی ارتش اسرائیل همسو است.[۲۱۲] در چنین شرایطی، کاربران سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) ممکن است بیشتر «عمل را بر عدم عمل ترجیح دهند»[۲۱۳] و این باعث می‌شود که بررسی دقیق اهداف به خطر بیفتد. توصیه‌های الگوریتمی ابتدا توسط تحلیلگران انسانی مرور می‌شوند تا تعیین شود که آیا برای بررسی بیشتر

---

[207] Asaf Lubin, 'The Reasonable Intelligence Agency' (2022) 47 Yale Journal of International Law 120.
[208] Sasso`li (n 126), 335-336.
[209] Boothby (n 8), 400.
[210] Boothby (n 8).
[211] Abraham (n 23).
[212] A source explained to the +972 Magazine and Local Call report: 'When a 3-year-old girl is killed in a home in Gaza, it' s because someone in the army decided it wasn' t a big deal for her to be killed - that it was a price worth paying in order to hit [another] target. We are not Hamas. These are not random rockets. Everything is intentional. We know exactly how much collateral damage there is in every home' : ibid.
[213] Neil Renic and Elke Schwarz, 'Inhuman-in-the-loop: ai-Targeting and the Erosion of Moral Restraint' (*Opinio Juris*, 19 December 2023) <https://opiniojuris.org/2023/12/19/inhuman-in-the-loop-ai-targeting-and-theerosion-of-moral-restraint/>.

تأیید شوند یا نه. با این‌حال، تحلیلگران انسانی ممکن است تحت فشار حجم زیاد داده‌ها و شرایط جنگی قرار داشته باشند و سوگیری‌های خود را در انتخاب اهداف تأیید کنند. این وضعیت باعث می‌شود فرماندهان نظامی نتوانند به‌طور کامل دقت توصیه‌ها را بررسی کنند و اعتماد کافی به سامانه نداشته باشند. بنابراین، نگرانی وجود دارد که آیا تصمیم‌گیرندگان قادر هستند مسئولیت تأیید دقت اهداف ارائه‌شده را حفظ کنند.

افزایش سرعت و تعداد توصیه‌ها مزیت تصمیم‌گیری سریع دارد، اما می‌تواند هزینه‌ای به شکل بررسی ناکافی اهداف و خطر آسیب به غیرنظامیان داشته باشد. اختصاص زمان بیشتر برای بررسی توصیه‌های الگوریتمی می‌تواند پایبندی به اقدامات احتیاطی را افزایش دهد، به‌ویژه با توجه به پیچیدگی هوش مصنوعی، سرعت بالای توصیه‌ها و فشار محیط جنگی.

نتیجه‌گیری

این مقاله یک بحث مقدماتی ارائه می‌دهد تا چالش‌هایی که ممکن است هنگام استفاده فرماندهان نظامی از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) در تصمیم‌گیری‌های هدفگیری ایجاد شود، ترسیم کند. فناوری به سرعت پیشرفت می‌کند و انتظار می‌رود ارتش‌ها سامانه‌هایی را توسعه دهند که تصمیم‌گیری سریع‌تر و کارآمدتری فراهم کنند.[۲۱۴] سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) می‌توانند حجم زیادی از داده‌ها را تحلیل کرده و توصیه‌هایی برای پشتیبانی از تصمیم‌گیری ارائه دهند.[۲۱۵] اگرچه این سامانه‌ها به‌طور مستقل «ماشه را نمی‌کشند» اما استفاده از آن‌ها نگرانی‌هایی ایجاد می‌کند، زیرا فرماندهان نظامی به توصیه‌های الگوریتمی متکی هستند.

استفاده مؤثر از سامانه‌های تشخیص چهره (FRT) به محیطی که این فناوری در آن به کار گرفته می‌شود وابسته است. در محیط‌های کنترل‌نشده، سامانه‌های تشخیص چهره (FRT) می‌توانند عدم دقت و احتمال مثبت‌های کاذب را افزایش دهند. قوانین بین‌المللی بشردوستانه هنوز مشخص نکرده‌اند که چگونه، چه زمانی و تا چه حد می‌توان از سامانه‌های تشخیص چهره (FRT) برای شناسایی یا تأیید افراد در هدفگیری استفاده کرد. تعامل بین سامانه‌های تشخیص چهره (FRT) و سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) نیاز به شفاف‌سازی بیشتری دارد تا مشخص شود تا چه حد می‌توان از آن‌ها برای پشتیبانی از تصمیم‌گیرندگان نظامی در عملیات هدفگیری استفاده کرد.

استفاده از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) نگرانی‌های حقوقی ویژه‌ای ایجاد می‌کند، زیرا توصیه‌های تولیدشده الگوریتمی با سرعت و مقیاس بالا ارائه می‌شوند و این می‌تواند بر قضاوت انسانی فرماندهان تأثیر بگذارد. همانطور که در مطالعه موردی لاوندر نشان داده شد، سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) می‌توانند برای شناسایی یا برچسب‌زنی اهداف بالقوه استفاده شوند، که این خود پرسش‌هایی درباره قضاوت و مسئولیت انسانی ایجاد می‌کند.

برای اطمینان از استفاده قانونی از سامانه‌های پشتیبانی تصمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) در درگیری‌ها، لازم است تأیید دقیق انجام شود تا مشخص شود اهداف پیشنهادی دقیق هستند و مطابق با حقوق بین‌الملل بشردوستانه (IHL) از حمله مستقیم محافظت می‌شوند. تمرکز بیش از حد بر تصمیم‌گیری سریع می‌تواند به آسیب به غیرنظامیان منجر شود. برای کاهش این خطر و رعایت تعهدات حقوق بین‌الملل بشردوستانه (IHL)، ممکن است لازم باشد که نقش سامانه‌های پشتیبانی

---

[214] Ekelhof (n 1).
[215] icrc, 'International Humanitarian Law and the Challenges of Contemporary Armed Conflicts' (Geneva, November 2019)

تصـمیم‌گیری مبتنی‌بر هوش مصـنوعی (AI-DSS) محدود به وظایف خاص شـود، اسـتفاده از آن در محیط‌های پرجمعیت غیرنظامیان محدود گردد و فرآیند تصـمیم‌گیری نظامی کند شـود تا تصـمیم‌گیرندگان زمان کافی برای ارزیابی کیفی اهداف داشته باشند.

گزارش‌های مربوط به لاوندر نشـان می‌دهد که شـفافیت در این زمینه کم اسـت و این مسـئله بر توانایی پژوهشـگران برای درک چگونگی اسـتفاده ارتش‌ها از سـامانه‌های پشـتیبانی تصـمیم‌گیری مبتنی‌بر هوش مصـنوعی (AI-DSS) در جنگ تأثیر می‌گذارد. عدم شـفافیت دولت‌ها درباره اسـتفاده از هوش مصـنوعی نگرانی‌هایی جدی درباره دسـترسـی به شـواهد برای تحقیقات و حفظ مسـئولیت در برابر نقض قوانین بین‌المللی ایجاد می‌کند و بر نظارت بر طراحی، توسـعه و اسـتفاده از این فناوری‌ها مانع ایجاد می‌کند. بنابراین، از دولت‌ها خواسـته می‌شـود که درباره اسـتفاده، سـیاسـت‌ها و مقررات سـامانه‌های پشـتیبانی تصـمیم‌گیری مبتنی‌بر هوش مصنوعی (AI-DSS) در درگیری‌های مسلحانه شفاف باشند.